

REPORT DOCUMENTATION PAGE				Form Approved OMB NO. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 27-09-2012		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 1-Mar-2005 - 31-Dec-2008	
4. TITLE AND SUBTITLE Final Technical Report Project W911NF-05-1-0067 Feature-Based Methods For Landmine Detection With Ground Penetrating Radar				5a. CONTRACT NUMBER W911NF-05-1-0067	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 654808	
				5d. PROJECT NUMBER 633606	
6. AUTHORS Paul D. Gader, Joseph N. Wilson and Jeremy Bolton				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Florida - Gainesville Office of Sponsored Program University Of Florida Gainesville, FL 32611 -				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211				10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) 47999-EV.8	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The subject research was performed at the University of Florida between December 2005 and December 2008. The research was performed to support the ability to detect landmines in an automated fashion using ground-penetrating radar (GPR) array sensors employed in systems being studied by NVESD. The work was concerned with discovering and evaluating i) different types of features that, when extracted from signals associated with GPR signals captured over regions of earth, can help one identify the presence or absence of landmines and					
15. SUBJECT TERMS Landmine detection, University of Florida, Pattern recognition, Image processing, multi-sensor fusion, classifier development					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Paul Gader
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 352-505-2551

Report Title

Final Technical Report
Project W911NF-05-1-0067
Feature-Based Methods For Landmine Detection
With Ground Penetrating Radar

ABSTRACT

The subject research was performed at the University of Florida between December 2005 and December 2008. The research was performed to support the ability to detect landmines in an automated fashion using ground-penetrating radar (GPR) array sensors employed in systems being studied by NVESD. The work was concerned with discovering and evaluating i) different types of features that, when extracted from signals associated with GPR signals captured over regions of earth, can help one identify the presence or absence of landmines and landmine-like objects; ii) algorithms and techniques that can employ these features to distinguish between landmines and non-mines; and iii) fuse the results of multiple discriminators to yield improved discrimination performance.

This document briefly reviews results of this research in each of these areas. Referenced papers are attached as appendices.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

09/27/2012	5.00	J. N. Wilson, P. D. Gader. Use of the Borda count for landmine discriminator fusion, Detection and Remediation Technologies for Mines and Minelike Targets XII. , Orlando, FL, USA. : ,
09/27/2012	6.00	H. Frigui, P. Gader. Detection and Discrimination of Land Mines based on Edge Histogram Descriptors and Fuzzy K-Nearest Neighbors, 2006 IEEE International Conference on Fuzzy Systems. 2006/07/01 00:00:00, Vancouver, BC, Canada. : ,

TOTAL: 2

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received

Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	Discipline
Arthur Kumarak	0.50	
Jeremy Bolton	0.50	
Ryan Busser	0.50	
Gyeongyong Heo	0.50	
Taylor Glenn	0.50	
Razia Mazhar	0.25	
Andres Mendez-Vazquez	0.50	
Bradley Mouring	0.10	
Ganesan Ramachandran	0.50	
Seniha Yuksel	0.50	
Alina Zare	0.10	
Xuping Zhang	0.50	
FTE Equivalent:	4.95	
Total Number:	12	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
Wen Lee	1.00
FTE Equivalent:	1.00
Total Number:	1

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	National Academy Member
Paul Gader	0.25	
Joseph Wilson	0.25	
Mark Schmalz	0.15	
FTE Equivalent:	0.65	
Total Number:	3	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

<u>NAME</u>
Ryan Busser
Taylor Glenn
Total Number: 2

Names of personnel receiving PHDs

<u>NAME</u>
Andres Mendez-Vazquez
Jeremy Bolton
Alina Zare
Total Number: 3

Names of other research staff

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Report.

Technology Transfer

Final Technical Report
Project W911NF-05-1-0067
Feature-Based Methods For Landmine Detection
With Ground Penetrating Radar

Prepared by
Paul D. Gader, Joseph N. Wilson and Jeremy Bolton
Computer and Information Science and Engineering Department
University of Florida

Note: This document was previously submitted with an incorrect project number listed (W909MY-05-D-0001/0002). The project title (Feature-Based Methods For Landmine Detection With Ground Penetrating Radar) on the previously submitted report was correct. We regret any difficulties this error may have caused. This note added on 6 April 2011.

Introduction

The subject research was performed at the University of Florida between December 2005 and December 2008. The research was performed to support the ability to detect landmines in an automated fashion using ground-penetrating radar (GPR) array sensors employed in systems being studied by NVESD. The work was concerned with discovering and evaluating i) different types of features that, when extracted from signals associated with GPR signals captured over regions of earth, can help one identify the presence or absence of landmines and landmine-like objects; ii) algorithms and techniques that can employ these features to distinguish between landmines and non-mines; and iii) fuse the results of multiple discriminators to yield improved discrimination performance.

This document briefly reviews results of this research in each of these areas. Referenced papers are attached as appendices.

Features

During the period of performance, we investigated a wide variety of features arising from GPR signals, however, those features can be broken into several broad categories:

1. Spectral features, characterizing properties of the energy frequency spectrum of the radar signal return.
2. Spatial edge features, characterizing the locations and local spatial organizations of instantaneous changes in the radar signal return.

3. Spatial region features, characterizing the locations and extents of spatially contiguous radar returns having similar properties.

In each of the following subsection, we briefly describe the results of research associated with each of these types of features.

Spectral Features

Within this first category of features, we worked together with Dominic Ho of the University of Missouri to identify spectral properties of radar signals that were suggestive of the presence of landmines. The GPR signals we process contain a wide variety of clutter objects such as rocks and roots, and they also display great soil heterogeneity. We identified frequency domain spectral features that improve the detection of weak-scattering plastic mines and to reduce the number of false alarms resulting from clutter in comparison to earlier algorithms. The motivation for this approach arose from the fact that landmine targets and clutter objects have different shapes and/or composition that yield different energy density spectrum (EDS) that may be exploited for discrimination. Although the same information is present in time-domain data, the frequency domain lets us remove the phase component and can reveal better spatial characteristics and often achieve greater robustness. The EDS (Ho, et al., 2008) is essentially a spatially averaged energy signature extracted from a normalized and signal-smoothed region surrounding a point to be extracted. The consistency of the landmine spectral characteristics was confirmed by data collected at several geographically diverse sites having different soil conditions and by the data produced from two completely different radar systems. Experimental results corroborated the effectiveness of the spectral features in improving landmine/clutter discrimination and the robustness of the EDS estimation method.

Spatial Edge Features

A variety of different spatial edge features were investigated during the course of this research.

The first edge features were exploited in a Hidden Markov Model (HMM) detector (Wilson, et al., 2007). The observations used by the HMM were positive and negative diagonal and antidiagonal edges found in the second derivative of the B-scans of the radar data.

Work reported by Frigui and Gader (Frigui, et al., 2006) describes a set of edge histogram features motivated by the MPEG compression standard, and the edge histogram descriptor which collects together a set of spatially organized edge features. These features are created by finding the edge orientation at each pixel to be inspected by applying four edge masks (horizontal, vertical, diagonal, antidiagonal) and finding the greatest of the responses that exceeds a required threshold. If the threshold is not exceeded, an anisotropic edge is reported. A histogram over these five responses is created for a block, or collection of spatially neighboring pixels. The EHD collects together the histograms over a number of spatially neighboring blocks to characterize the edges associated with a region of earth corresponding to a collection of neighboring B-scans.

Spatial Region Features

In addition to the edge features, a number of spatial region features of the time-domain GPR signal have been employed in attempting to identify landmines. The features found to be of greatest utility (Wilson, et al., 2007) can be roughly comprise a number of energy region characteristics. After finding connected components of high energy (as identified by exceeding the Otsu threshold in depth-bin whitened data), we calculate the following region properties: eccentricity, solidity, area-to-filled-area ratio, compactness.

Discrimination Algorithms

The features identified in the program were employed in several different discrimination algorithms to attempt to yield mine confidence values giving high detection probability with correspondingly low false alarm rates. This section briefly describes the algorithms developed and employed in this investigation.

Hidden Markov Model

The HMM algorithm (Wilson, et al., 2007) uses observation sequences that are the diagonal and antidiagonal edge features discussed above. An HMM begins execution in what is referred to as its initial state. Thenceforth, it determines the most likely state given the previous state and the current observation. The HMM is trained using data that identifies those points in the observation sequence associated with the presence of landmines. To identify the existence of a mine, one finds the likelihood of being in a landmine state at a given time and report this as the mine confidence value.

Feed-forward Ordered-Weighted Average

The Feed-forward Ordered-Weighted Average (FOWA) algorithm employs depth-bin specific spatial region features (those identified above) as input to a multilayer perceptron having an input layer whose feature values are bin-sorted values, thus, their ordering in depth is lost, however, their ordering in the feature space is preserved. In addition to the typical MLP training, we have modified the system to train in such a way as to maximize the area under the receiver-operating characteristics (ROC) curve (Lee, et al., 2007).

Edge Histogram Detector

The Edge Histogram Detector (Frigui, et al., 2006) clusters the EHD features identified above and using a possibilistic K-nearest neighbor approach to identify the most likely class associated with a region of earth under inspection.

Spectral Confidence Feature

A spectral confidence feature can be formed by calculating the EDS of a region of earth under inspection, then applying a matched filter developed by analysis of weak-scattering low-metal

landmines in order to attempt to associate high confidence values with these difficult-to-find mines (Ho, et al., 2008).

Summary of Algorithms

These algorithms were employed over a wide range of targets in diverse environments with both naturally-occurring and emplaced clutter objects to identify their performance characteristics. It was found that they provide contrasting benefits in different environments with different target sets.

Fusion of Decision Statistics

As noted above, each of the discrimination algorithms exhibits differing characteristics with respect to their performance on particular landmine classes and in specific environments. We studied and developed methods to fuse their results for improved discrimination performance.

Choquet Measure Fusion

One fusion method we studied was to use the Choquet integral of differing detector outputs to yield an improved result. One of the methods that yielded good results was to use minimum classification error training (Mendez-Vazquez, et al., 2008).

Rank-based Fusion

Another method employed in this work was to use rank-based fusion, which normalizes algorithm decision statistics by their rank in a training set, then combines linearly-weighted ranks to yield a fusion result (Frigui, et al., 2009).

Summary

A variety of features, discriminators, and fusion methods for detecting the presence of landmines in GPR signals with high probability and low false alarm rates were studied and reported on. In addition to publishing the results of our work, we presented these results at Algorithm Working Group Meetings, conveyed algorithms and methods to military contractors, and communicated methods and results to sponsor representatives and other interested parties. We have attached relevant documents as Appendices

Bibliography and Appendices

Frigui H. [et al.] An evaluation of several fusion algorithms for anti-tank landmine detection and discrimination [Journal] // Information Fusion. - [s.l.] : Elsevier, 2009.

Frigui H. and Gader P.D. Detection and Discrimination of Land mines based on Edge [Conference] // International Conference on Fuzzy Systems. - Vancouver, BC : IEEE, 2006. - pp. 1494-1499.

Ho K.C. [et al.] An Investigation of Using the Spectral Characteristics From Ground Penetrating Radar for Landmine/Clutter Discrimination [Journal] // IEEE Trans. Geoscience and Remote Sensing. - April 2008. - 4 : Vol. 46. - pp. 1177-1191.

Lee W-H., Gader P.D. and Wilson J.N. Optimizing the Area Under a Receiver Operating Characteristic Curve With Application to Landmine Detection [Journal] // IEEE Trans. Geoscience and Remote Sensing. - February 2007. - 2 : Vol. 45. - pp. 389-397.

Mendez-Vazquez A. [et al.] Minimum Classification Error Training for Choquet Integrals With Applications to Landmine Detection [Journal] // IEEE Trans. Fuzzy Systems. - February 2008. - 1 : Vol. 16. - pp. 225-238.

Wilson J.N. [et al.] A Large-Scale Systematic Evaluation of Algorithms Using Ground-Penetrating Radar for Landmine Detection and Discrimination [Journal] // IEEE Trans. Geoscience and Remote Sensing. - August 2007. - 8 : Vol. 45. - pp. 2560-2572.

Wilson J.N. and Gader P.D. Use of the Borda Count for Landmine Discriminator Fusion [Conference] // Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets. - Orlando, FL : SPIE, 2007. - Vol. 6553. - pp. 1112-1120.



An evaluation of several fusion algorithms for anti-tank landmine detection and discrimination

Hichem Frigui^{a,*}, Lijun Zhang^a, Paul Gader^b, Joseph N. Wilson^b, K.C. Ho^c, Andres Mendez-Vazquez^b

^aCECS Dept., University of Louisville, Louisville, KY, USA

^bCISE Dept., University of Florida, Gainesville, FL, USA

^cECE Dept., University of Missouri, Columbia, MO, USA

ARTICLE INFO

Article history:

Received 16 January 2008

Received in revised form 1 October 2008

Accepted 30 October 2009

Available online 10 November 2009

Keywords:

Landmine detection

Bayesian fusion

Context-dependent fusion

Dempster–Shafer

Borda count

Decision template

Fuzzy integral

ABSTRACT

Many algorithms have been proposed for detecting anti-tank landmines and discriminating between mines and clutter objects using data generated by a ground penetrating radar (GPR) sensor. Our extensive testing of some of these algorithms has indicated that their performances are strongly dependent upon a variety of factors that are correlated with geographical and environmental conditions. It is typically the case that one algorithm may perform well in one setting and not so well in another. Thus, fusion methods that take advantage of the stronger algorithms for a given setting without suffering from the effects of weaker algorithms in the same setting are needed to improve the robustness of the detection system. In this paper, we discuss, test, and compare seven different fusion methods: Bayesian, distance-based, Dempster–Shafer, Borda count, decision template, Choquet integral, and context-dependent fusion. We present the results of a cross validation experiment that uses a diverse data set together with results of eight detection and discrimination algorithms. These algorithms are the top ranked algorithms after extensive testing. The data set was acquired from multiple collections from four outdoor sites at different locations using the NIITEK GPR system. This collection covers over 41,807 m² of ground and includes 1593 anti-tank mine encounters.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

It is estimated that over 100 million landmines are buried in over 80 countries and that 26,000 people a year are killed or maimed by a landmine [1]. Detection and removal of landmines is a significant research problem [2–5]. The research problem for data analysis is to determine how reliably landmines can be detected and distinguished from other subterranean objects using sensor data. Difficulties arise from the variability of landmine types, soil and weather conditions, terrains, and so on. Traditional fielded approaches use metal detectors. Unfortunately, many landmines contain little metal. Ground penetrating radar (GPR) offers the promise of detecting landmines with little metal. Although several approaches to detecting landmines and discriminating landmines from clutter using GPR have been investigated [6–15], acceptable results have been elusive [16–18]. Although systems often achieve high detection rates, it is difficult to achieve the required low false alarm rates. Moreover, algorithm performance can vary significantly. Therefore, fusion methods that take advantage of the strengths of individual algorithms, overcome their

weaknesses, and achieve a higher accuracy than any individual algorithm are needed.

Multi-classifier, multi-algorithm, and multi-sensor fusion are critical components in landmine detection. Buried objects interact with the soil and any potential covering of the soil (such as a road surface). Physical properties of soil can vary significantly in small areas. For example, soil can be a heterogeneous mixture of soil types layered with a thin layer of top soil covering clay or asphalt covering gravel covering soil. Soil can have significantly varying density in a small region [19]. Roots of vegetation hold water. Rain or snow lead to variable moisture in the soil. Minerals can significantly affect the radar propagation through soil. In addition, the mine case can interact with different soils in different ways. For example, plastic casings have very similar electrical properties as soils under some conditions. Wood casings can absorb moisture. All these factors can have significant effects on GPR data and are generally unknown to an autonomous algorithm due to the wide variability over a small range. The implication for autonomous detection is that different types of algorithms are useful for different conditions. These different algorithms must use different signal conditioning, or Preprocessing, and feature extraction.

The objective of this paper is to present results of evaluating eight different anti-tank landmine discrimination algorithms and

* Corresponding author.

E-mail address: h.frigui@louisville.edu (H. Frigui).

the fusion of these algorithms using seven different methods. The generality, computational cost, and interpretability of the fusion methods is analyzed using a cross validation experiment that uses a diverse data set acquired from four outdoor test sites at different geographic locations. This collection covers over 41,807 m² of ground and includes 1593 anti-tank mine encounters. This collection contains multiple sub-collections taken at different times of the year and at very different locations in the United States as well as in Europe. Therefore, the experimental results, although not completely independent of mine type, soil conditions, etc., are probably at least as independent as any published results.

Section 2 describes the GPR data, Preprocessing, and prescreening. Section 3 outlines the distinct anti-tank landmine discrimination algorithms. Section 4 discusses the seven methods for fusing discrimination algorithms. Experimental results and analyses are presented in Section 5. Section 6 concludes.

2. Data preprocessing and prescreening

In this section, we briefly describe the GPR data, Preprocessing steps, and Prescreening. More detailed descriptions are in [20,21].

2.1. GPR data

The input data consist of a sequence of raw GPR measurements collected by a vehicle-mounted GPR array [22] (see Fig. 1a). The GPR collects 24 channels of data. Adjacent channels are spaced approximately 5 cm apart in the cross-track direction, and sequences (or scans) are taken at approximately 5 cm down-track intervals. The system uses an antenna that generates a wide-band pulse from 200 MHz to 7 GHz. Each A-scan, that is, the measured waveform collected in one channel at one down-track position, contains 416 GPR time samples, each corresponding to roughly 8 ps. We often refer to the time index as depth although, since the radar wave travels through different media, this index does not represent a uniform sampling of depth. Thus, we model GPR input data as a three-dimensional matrix of sample values, $S(z, x, y)$, $z = 1, \dots, 416$; $x = 1, \dots, 24$; $y = 1, \dots, N_s$, where N_s is the total number of collected scans, and the indices z , x , and y represent depth, cross-track position, and down-track positions respectively. GPR input data is illustrated in Fig. 1b.

Fig. 2 displays down-track B-scans (sequences of A-scans from a single channel) and cross-track B-scans (sequences of A-scans from a single scan). The surveyed object position is highlighted in each figure.

2.2. Data preprocessing

Preprocessing is an important step to enhance the mine signatures. The algorithm first identifies the ground bounce location as the global maximum of the signal and aligns the A-scans using these maxima. This alignment is necessary because the system cannot maintain the radar antenna at a fixed distance above the ground. The early time samples of each signal, up to few samples beyond the ground bounce are discarded. The remaining samples are divided into N depth bins which will be processed independently. The reason for this segmentation is to compensate for the high contrast between responses from deeply buried and shallow anomalies.

2.3. Anomaly detection

Our algorithm applies a Prescreener to reduce the volume of GPR data to be inspected. The Prescreener identifies distinct alarm locations in the data. It was designed to provide a high probability of detection so that more computationally intensive discrimination processing can be performed. False alarms are alarms that do not correspond to mines. The objective of the feature-based detection algorithms and their fusion is to distinguish between Prescreener alarms corresponding to landmines from false alarms. We use the Duke University NUKEv6 Prescreener, a variant of the least mean squares (LMS) Prescreener [20]. A version of this Prescreener was implemented in real-time in the system in Fig. 1. This Prescreener is applied to the energy at each depth bin, and assigns a confidence value to each point in the cross-track, down-track plane based on its contrast with a neighboring region. The cross-track x_s , and down-track y_s positions of the centers of algorithmically determined mine-like components are reported as alarm positions for further processing.

3. Discrimination algorithms

Generally, automated landmine discrimination algorithms consist of three phases: Preprocessing, feature extraction, and confidence assignment. Preprocessing performs tasks such as normalizing data, correcting for variations in height and speed, and removing stationary effects due to the system response. Previous methods include wavelets and Kalman filters [23,24], subspace methods and polynomial matching [25], and subtracting optimally shifted and scaled reference vectors [26]. Feature extraction reduces the Preprocessed data to a lower-dimensional, salient set of values that represent the data. The principal component

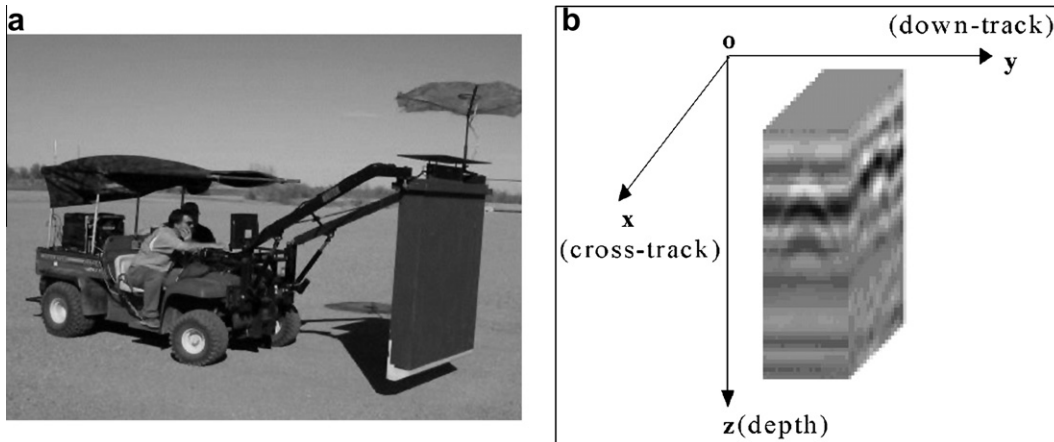


Fig. 1. GPR data collection: (a) NIITEK vehicle-mounted GPR system; and (b) an example of GPR scans.

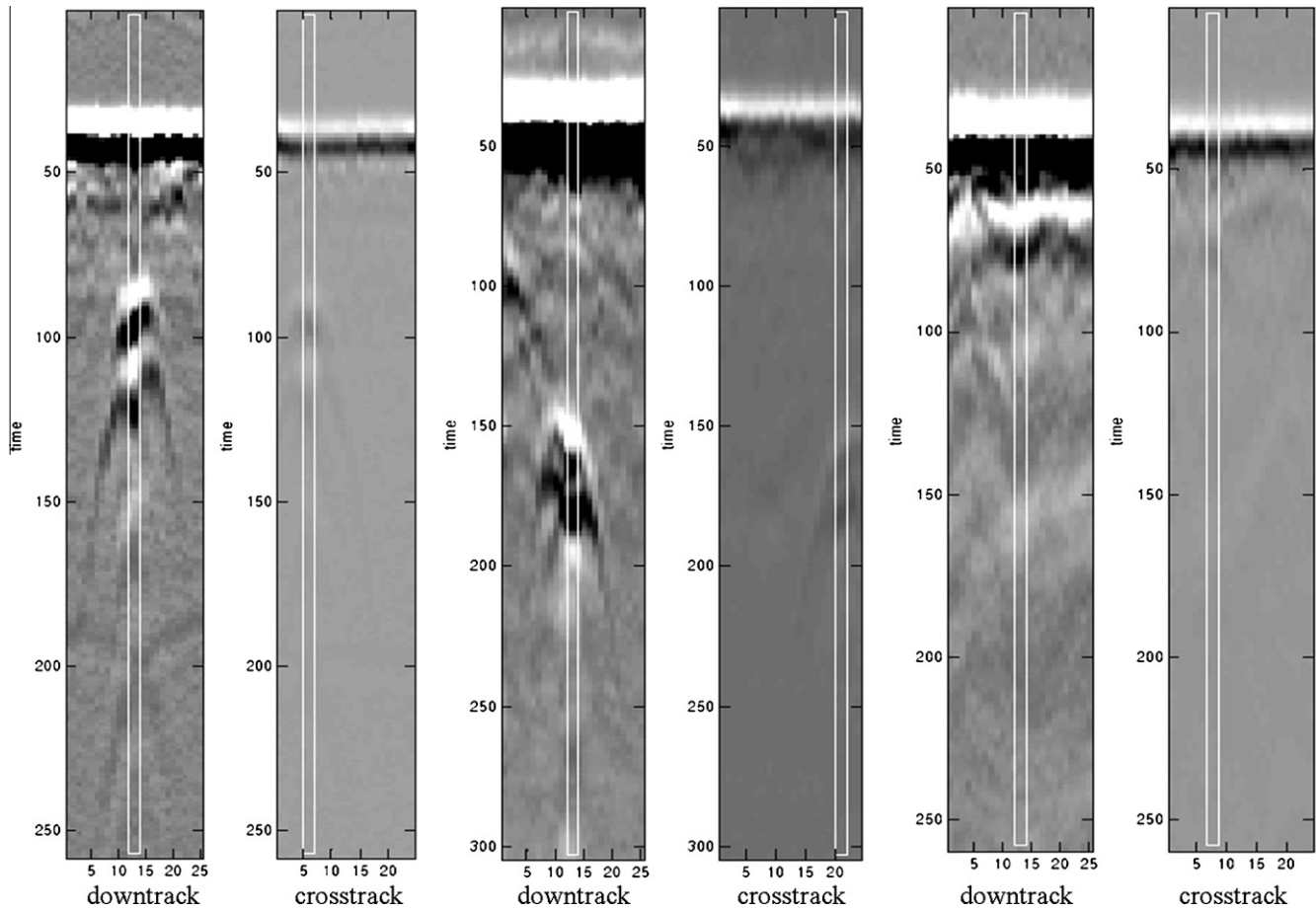


Fig. 2. NIITEK Radar down-track and cross-track B-scans pairs for three alarms.

transform is a common feature extraction tool [27], as are wavelets [23], image processing based differentiation [6], and Hough and Radon transforms [4]. Confidence assignment can be performed using methods such as Bayesian [4], hidden Markov Models [6,28], fuzzy logic [5], rules and order statistics [21], neural networks, or nearest neighbor classifiers [7].

Here we consider seven specific algorithms of distinct character. These algorithms have performed well in extensive field testing, and are being considered for real-time implementation in handheld and vehicle-mounted GPR systems. These algorithms are highlighted in the following sections.

3.1. HMM detector

The HMM algorithm [6,28] treats the down-track dimension as the time variable and produces a mine confidence at positions, (x, y) , on the surface being traversed. A sequence of observation vectors is produced for each down-track point and depth. These observation vectors encode the degree to which edges occur in the diagonal and anti-diagonal directions. In particular, for every point (x_s, y_s) , the strengths for the positive/negative diagonal/anti-diagonal edges is computed. The observation vector at a point (x_s, y_s) consists of a set of four features that encode the maximum edge magnitude over multiple depth values around (x_s, y_s) . The HMM algorithm has a background and a mine model. Each model produces a probability. The probability produced by the mine (background) model is an estimate of the probability of the observation sequence given that there is a mine (background) present. The log of the ratio of the probabilities is the confidence.

3.2. EHD detector

This algorithm uses translation invariant features based on the edge histogram descriptor (EHD) of the 3-D GPR signatures, and a possibilistic K-Nearest Neighbors (K-NN) rule for confidence assignment [29]. The EHD captures the signature's texture. Specifically, each 3-D signature is divided into sub-signatures, and the local edge distribution for each sub-signature is represented by a histogram. To generate the histogram, local edges are categorized into five types: vertical, horizontal, diagonal (45° rising), anti-diagonal (45° falling), and non-edges. A set of alarms with known ground truth is used to train the decision-making process. These labeled alarms are clustered to identify a small number of representatives that capture signature variations due to differing environmental conditions and mine types, etc.

3.3. SPECT detector

This detector aims at capturing the characteristics of a target in the frequency domain using the energy density spectrum (EDS). It extracts the spectral correlation feature (SCF) which is computed using similarity to mine prototypes [30]. The EDS is estimated using three main steps: Preprocessing, whitening, and averaging. After alignment, Preprocessing removes the data above and near the ground surface to avoid an EDS that is dominated by the ground response. The whitening step equalizes the background spectrum so the estimated EDS reflects the actual spectral characteristics of an alarm. Averaging reduces the variance in the EDS.

3.4. GEOM detector

This algorithm computes geometric features in multiple, whitened depth bins which are two-dimensional images with cross-track and down-track axes. The features are inputs to a Feed-forward Ordered-Weighted-Average (FOWA) network [31] that is trained to maximize the area under the Receiver Operating Characteristic (ROC) curve [32]. The features used are compactness, eccentricity, solidity, and area to filled area ratio. These features are based on the observation that the whitened energy for mines often has a compact, solid, and circular shape whereas non-mine-like objects produce an irregular shape.

3.5. TFCM detector

The Texture Feature Classification Method (TFCM) detector [33] is a three-dimensional extension of the algorithm by Horng [34]. The algorithm transforms a block of GPR data into a block of integer codes. The code at each point in a block is generated by considering several differences in GPR intensity values over a $3 \times 3 \times 3$ window centered at the point. The differences are thresholded producing a string of zeros and ones, which are then mapped to the integer codes, the details of which are described in the references. Statistical textures features, such as entropy, variance, and co-occurrence, are then computed on the blocks of codes and transformed into feature vectors. Relevance Vector Machines (RVMs) use the features to produce a confidence that an alarm represents a landmine.

3.6. GMRF detector

The Gaussian-Markov Random Field (GMRF) detector [35] is based on a transmission line model of the time-domain GPR response to the subsurface. The model represents the GPR as a sequence of dielectric discontinuities. Each discontinuity is parameterized by a location and a gain parameter. These parameters are characterized statistically using a Gaussian-Markov Random Field. A generalized likelihood ratio test is then used to assign a confidence that an alarm represents an anti-tank landmine.

3.7. GFIT detector

The Gaussian Fit (GFIT) detector [36] calculates the parameters of a Gaussian pulse which best fits the spatial energy distribution of target responses to GPR. The output features are the goodness of fit, the pulse width, and pulse gain. More specifically, the spatial shape of the summed energy from a cross-track scan is compared to the shape of a Gaussian pulse. If x represents position in down-track scans, and E represents the energy, we find the σ , x_0 , α to minimize the root mean square error (RMSE) between $E(x)$ and $f(x) = \alpha * \exp(-(x_0 - x)/\sigma^2)$. The output features are then $\sqrt{\sum_x (E(x) - f(x))^2}$, σ , x_0 , and α .

The above discrimination algorithms were developed by researchers at the Universities of Missouri, Louisville, Florida, as well as Duke University. They are independently developed and have many differences in GPR Preprocessing and normalization, feature extraction, and classification methodologies. Since the descriptions of almost all of these algorithms are contained in detail in the references, and take many pages to describe in detail, they cannot be described in detail here. However, in feature extraction alone one can see many differences. The anomaly detector simply looks for locations that are different from the background. It uses masks oriented in the C-scan direction. The HMM detector looks at variable length sequences of edges. The EHD detector looks at fixed length representations of edges. All three previous algorithms used the down-track and cross-track time-domain GPR.

The SPECT detector looks at features in the frequency domain. The GEOM detector calculates feature based on geometric shape in C-scans. The TFCM detector looks for texture features in three-dimensional blocks of time domain data, GMRF, and the GFIT detector looks at energy in the cross-track direction. Thus, in the feature extraction process alone, one can see that these algorithms vary widely in the focus and processing.

Despite all of the above differences, one cannot assume that these algorithms are statistically independent. In fact, we know that some of them could be highly correlated. For instance, both the EHD and the HMM detectors could assign low confidence values to alarms with weak edges. The fusion algorithms that we are considering (described in the next section) address the independence issue to various degrees. For instance, the Bayes fusion and the Mahalanobis distance fusion do not make the independence assumption and use full covariance matrices to normalize and decorrelate the detectors outputs. Similarly, the Choquet integral considers all possible subsets of detectors and promotes sparsity. Thus, it will tend to identify the smallest subset of uncorrelated detectors. Other fusion methods do not consider the detectors dependency at all. One of the goals of this experiment is to compare these fusion methods with respect to this dependency issue.

4. Combination of multiple classifiers

4.1. Background

For complex detection and classification problems involving data with large intra-class variations and noisy inputs, perfect solutions are difficult to achieve, and no single source of information can provide a satisfactory solution. As a result, combination of multiple classifiers (or multiple experts) is playing an increasing role in solving these complex pattern recognition problems, and has proven to be viable alternative to using a single classifier. Classifier combination is mostly heuristic and is based on the idea that classifiers with different methodologies or different features can have complementary information. Thus, if these classifiers cooperate, group decisions should be able to take advantages of the strengths of the individual classifiers, overcome their weaknesses, and achieve a higher accuracy than any individual's.

Methods for combining multiple classifiers can be classified into two main categories: classifier selection and classifier fusion. Classifier selection methods assume that the classifiers are complementary, and that their expertise varies according to the different areas of the feature space. For a given test sample, these methods attempt to predict which classifiers are more likely to be correct. Some of these methods consider the output of only one classifier to make the final decision [37]. Others, combine the output of multiple "local expert" classifiers [38]. Classifier fusion methods assume that the classifiers are competitive and are equally experienced over the entire feature space. For a given test sample, the individual classifiers are applied in parallel, and their outputs are combined in some manner to take a group decision.

Over the past few years, a variety of schemes have been proposed for combining multiple classifiers. The most representative approaches include majority vote [39], Borda count [40], average [41], weighted average [42], Bayesian [43], and probabilistic [44]. Most of the above approaches assume that the classifier decisions are independent. However, in practice, the outputs of multiple classifiers are usually highly correlated. Therefore, in addition to assigning fusion weights to the individual classifiers, it is desirable to assign weights to subsets of classifiers to take into account the interaction between them. Fusion methods based on the fuzzy integral [45,46] and Dempster-Shafer theory [47,48] have this desirable property.

Another way to categorize classifier combination methods is based on the way they select or assign weights to the individual classifiers. Some methods are global and assign a degree of worthiness, that is averaged over the entire training data, to each classifier. Other methods are local and adapt the classifiers' worthiness to different data subspaces. Intuitively, the use of data-dependent weights, when learned properly, provides higher classification accuracy. This approach requires partitioning the input samples into regions during the training phase. The partition can be defined from the space of individual classifier decisions [49], according to which classifiers agree with each other [40], or by features of the input space [50,51]. Then, the best classifier for each region is identified and is designated as the expert for this region [52]. Conversely, the partitioning can be defined such that each classifier is an expert in one region [37]. This approach may be more efficient, however, its implementation is not trivial. In the classification phase, the region of an unknown sample is identified, and the output of the classifier responsible for this region is used to make the final decision. Data partition and classifier selection could also be made dynamic during the testing phase [53,54]. In this case, the accuracy of each classifier (with respect to the training samples) is estimated in local regions of the feature space in the vicinity of the test sample. The most accurate classifier is selected to classify the test sample.

4.2. Notation

Let $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L$ denote the L algorithms to be fused, and let w_1, \dots, w_C denote the C classes. Each algorithm, \mathcal{D}_i , extracts a set of features, F_i , and assigns a confidence value y_i to each of the C classes. In the proposed landmine application, we have $L = 8$, where $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_8$ correspond to the prescreener (NUKEv6), EHD, HMM, Spect, Geom, TFCM, GFIT, and GMRF algorithms respectively. We also have $C = 2$ where w_1 denotes the mine class and w_2 denotes the clutter class. We note that the prescreener is not a feature-based algorithm, and thus, it does not generate a set of features (i.e., no F_1).

4.3. Bayesian-based fusion

Bayesian data fusion [55] is based on Bayesian decision theory which is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the trade-offs between various classification decisions using probability and the costs that accompany such decisions. Bayesian data fusion has been studied extensively in the literature (e.g. [55–57]). This approach has the advantage of being able to incorporate *a priori* knowledge about the likelihood of the hypothesis being tested, and when empirical data are not available, it is possible to use subjective estimates of the prior probabilities. Moreover, from a statistical point of view, the use of Bayes rule should provide the optimal decision. Unfortunately, the proper use of Bayes requires the joint probability density functions to be known. This information is usually not available and may be difficult to estimate from the data. Other disadvantages of the Bayesian approach include complexities when dealing with multiple potential hypotheses and multiple conditionally dependent events, and the inability to account for general uncertainty [56,57]. Thus, Bayesian data fusion is best suited to applications where prior parameters are available, there is no need to represent ignorance, and where conditional dependencies can be easily modeled through probabilistic representation.

Bayesian fusion has been applied to target identification [58], image analysis [59], and many other applications [55]. It has also been applied to the problem of anti-personnel landmine detection [60,61], and the results were compared to other fusion methods. In

[60], only synthetic data were used, and in [61] a very small data set was used. Thus, the results were not conclusive.

Let \mathbf{v} represents the output of all L algorithms to be fused, i.e., $\mathbf{v} = [y_1, y_2, \dots, y_L]$. Within the Bayesian framework, \mathbf{v} is considered a random variable with a distribution that depends on the state of nature. Using Bayes formula, we first compute the posterior probability using

$$p(w_i|\mathbf{v}) = \frac{p(\mathbf{v}|w_i)p(w_i)}{p(\mathbf{v})}. \quad (1)$$

Then, \mathbf{v} is assigned to the class with maximum posterior probability, i.e.,

$$\mathbf{v} \in w_j \quad \text{if} \quad p(w_j|\mathbf{v}) = \max_{i=1 \dots K} p(w_i|\mathbf{v}). \quad (2)$$

In (1), $p(w_i)$ is the prior probability of class i and $p(\mathbf{v}|w_i)$ is the class conditional density. The prior $p(w_i)$ is usually provided by an expert, or estimated using the relative proportions of training data from each class. Similarly, $p(\mathbf{v}|w_i)$ can be estimated from the training data.

Our data consist of multiple subsets of mines/clutter signatures collected with the same hardware at different times and under different conditions. Moreover, many mines are of the same type and/or buried at the same depth. Thus, it is reasonable to assume that the detectors will assign confidence values that are consistent with these conditions, and that the confidence values of all detectors tend to form clusters in the confidence space. Consequently, we model $p(\mathbf{v}|w_i)$ by a mixture of M Gaussian distributions, i.e.,

$$p(\mathbf{v}|w_i) = \sum_{k=1}^M p(\mathbf{v}|w_{ki})P(w_{ki}), \quad (3)$$

where each $p(\mathbf{v}|w_{ki})$ is a multi-variate Gaussian. In general, we have $P(w_2) \gg P(w_1)$. However, the risk associated with missing a mine is much higher than the risk associated with detecting a false alarm. Since we cannot quantify the risks, and the priors can change from one site to another and depend on the settings of the prescreener, we simply assume that these two factors cancel each other, and let $P(w_1) = P(w_2)$.

In our experiments, we let \mathbf{v} include the output of the seven detection algorithms and the prescreener, i.e., $\mathbf{v} = [y_1, y_2, \dots, y_8]$. The means μ_{ki} , covariance matrices Σ_{ki} , number of components M , and the mixing coefficients $P(w_{ki})$ for the M components of class i are learned from the training data using the competitive agglomeration clustering algorithm [62]. Instead of using (2) to label the test data, we assign a soft confidence value using

$$\text{Conf}_B = P(w_1|\mathbf{v}). \quad (4)$$

4.4. Mahalanobis distance-based fusion

The Mahalanobis distance-based approach (MD) is a variation of the Bayesian approach [63]. It models the distribution of \mathbf{v} in each class i , by a multi-variate Gaussian and therefore represents the eccentricity of the mine and clutter distributions. The Mahalanobis distances to the mine and clutter classes of a test alarm \mathbf{v} are computed by

$$D_X = (\mathbf{v} - \mu_X)^T \Sigma_X^{-1} (\mathbf{v} - \mu_X), \quad (5)$$

where $X = M$ and $X = C$ denote the mine and clutter classes, respectively. The fusion confidence is the weighted difference between the distances:

$$\text{Conf}_{MD} = -D_M + \alpha D_C. \quad (6)$$

The value α in (6) provides a means of controlling the contribution of the distance to the clutter class to the fusion confidence. It is

computed from the training data to minimize the average false alarm rate over the range of probability of detection from 92% to 96% [63]. This range was chosen since our long term goal is in probabilities of detection around 95% and this interval contains that range. Based on our experience the mines with confidence so low that they are within the last 4% of the mines detected tend to be lucky detects, i.e. they do not really produce useful signatures and therefore should not be included in the optimization. This is why the range is not symmetric around 95%.

The use of Mahalanobis distance has the advantage of normalizing the features and removing their correlation before fusing them. This is reflected by the use of the covariance matrix in the distances (5). Furthermore, the generation of confidences using (6) is based on the theoretically sound likelihood ratio when \mathbf{v} is assumed to be Gaussian and when $\alpha = 1$ [64].

4.5. Dempster–Shafer based Fusion

Dempster–Shafer (DS) is a mathematical theory of evidence for representing uncertain knowledge [65,66]. In a finite discrete space, DS can be interpreted as a generalization of probability theory where probabilities are assigned to sets as opposed to mutually exclusive singletons. In DS, evidence can be associated with multiple possible events, e.g., sets of events. As a result, evidence in DS can be meaningful at a higher level of abstraction without having to resort to assumptions about the events.

DS fusion was applied to handwriting recognition [67], decision making [68], face detection [69], landmine detection [60,61,48], and more [55,70]. One important feature of DS is the ability to cope with varying levels of precision regarding the information with no further assumptions needed to represent the information. It also allows for direct representation of uncertainty of system responses. However, DS fails to give an acceptable solution to fusion problems with significant conflict [71,72]. Consequently, many researchers developed modified Dempster rules to represent the degree of conflict [70].

DS and Bayesian theories have been studied and compared extensively [73,57,74]. Both theories have initial requirements. DS theory requires masses to be assigned to alternatives in a meaningful way, including the unknown state; whereas Bayes theory requires prior probabilities. In general, the results of both methods may be comparable, but the implementations may require different amounts of effort and information. Thus, selecting one approach over the other usually depends on the extent to which prior information is available.

Let $\Theta = \{\theta_1, \dots, \theta_k\}$ be a finite set of possible hypotheses, also referred to as the frame of discernment. The basic belief assignment function, m , a primitive of evidence theory, assigns a value in $[0, 1]$ to every subset A of Θ and satisfies

$$m(\phi) = 0, \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1. \quad (7)$$

$m(A)$ is the belief that supports A , but makes no additional claims about any of the subsets of A . Two basic belief functions m_1 and m_2 can be combined to obtain the belief mass committed to $C \subset \Theta$ as follows [66],

$$m(C) = m_1(C) \oplus m_2(C) = \frac{\sum_{j,k, A_j \cap B_k = C} m_1(A_j) m_2(B_k)}{1 - \sum_{j,k, A_j \cap B_k = \phi} m_1(A_j) m_2(B_k)}, \quad C \neq \phi. \quad (8)$$

This combination rule is extended to several belief functions by repeating the rule for new belief functions. The denominator in (8) is a normalizing factor, which intuitively measures how much m_1 and m_2 are conflicting. This normalization has the effect of completely ignoring conflict and causing any belief mass associated

with conflict to the null set [71]. Consequently, in the case of a significant conflict, this normalization can yield counterintuitive results. Fortunately, for the application under consideration, alarms with strongly conflicting evidence are unlikely. This is because all of the discrimination algorithms considered here use data from the same sensor (GPR) and try to identify signatures that have a consistent shape.

In some applications, we have prior knowledge about reliability of the sources. In this case, we can assign them weights before combining their belief functions, resulting in a weighted Dempster–Shafer fusion rule:

$$m(C) = m_1 \oplus m_2(C) = \frac{\sum_{j,k, A_j \cap B_k = C} w_1 m_1(A_j) w_2 m_2(B_k)}{1 - \sum_{j,k, A_j \cap B_k = \phi} w_1 m_1(A_j) w_2 m_2(B_k)}, \quad C \neq \phi. \quad (9)$$

Since we have classes mine (M) and clutter (C), we build the frame of discernment as $\Theta = \{\emptyset, \{M\}, \{C\}, \{M, C\}\}$. For each individual algorithm i , we associate a basic belief function m^i such that

$$m_i(\{M\}) = p_i^m, \quad m_i(\{C\}) = p_i^c, \quad \text{and} \quad m_i(\{M, C\}) = 1 - p_i^m - p_i^c, \quad (10)$$

where p_i^m and p_i^c are the confidences in the mine and clutter classes generated by algorithm i . These values are computed from the algorithms' confidence values as follows. First, we separate the training mine alarms from the clutter alarms and, for each algorithm i , we compute the cumulative probability distribution of each class, G_i^m and G_i^c . Then, we compute $p_i^m = G_i^m(y_i)$ and $p_i^c = 1 - G_i^c(y_i)$. Since p_i^m and p_i^c are computed independently using the training data of each class, they are not constrained to sum to 1.

The fusion of the eight algorithms is performed by combining their basic belief functions using (8) or (9). In the latter case, the weights are obtained from training data based on individual algorithm performance. The final mine confidence is

$$Conf_{DS} = m(\{M\}) - m(\{C\}) + K, \quad (11)$$

where K is a constant used to ensure that $Conf_{DS} \geq 0$ for all test samples.

4.6. Decision template fusion

Decision template (DT) is a fusion scheme that combines classifier outputs by comparing them to a characteristic template for each class [75]. DT fusion uses all classifier outputs to calculate the final support for each class, which is in sharp contrast to most other fusion methods which use only the support for that particular class to make their decision. The DT approach treats the classifier outputs as input to a second-level classifier in some intermediate feature space, and designs a new classifier for the second (combination) level.

DT fusion is computationally simple and does not rely on questionable assumptions. However, it does not consider the possible correlation among the individual classifiers. Moreover, its performance may depend on the distribution of the classifiers' output which can affect the similarity measure in the intermediate feature space. DT fusion has been applied to various areas such as time series classification [76], biometrics [77], and intrusion detection [78], and compared with many other fusion techniques. The results are in general inconclusive, which confirms that there is no fusion method that outperforms all others in all applications.

Let $d_{ij}(\mathbf{x}) \in [0, 1]$ represent the degree of support given by algorithm i to the hypothesis that \mathbf{x} comes from class w_j (e.g. the posterior probability $P(w_j|\mathbf{x})$). The outputs of all classifiers are organized in a *decision profile* matrix $\mathcal{DP}(\mathbf{x})$. The value in row i and column j of the decision profile matrix is $d_{ij}(\mathbf{x})$ [75]. Using

$\mathcal{DP}(\mathbf{x})$ as an *intermediate features space*, one can build a minimum-error classifier by replacing the problem of estimating $P(w_i|\mathbf{x})$ with one of estimating $P(w_i|\mathcal{DP}(\mathbf{x}))$. Thus, the initial feature space with n features, \mathbb{R}^n , is transformed into a new space with $L \times C$ features.

Training consists of calculating one DT per class using the training data. Let Z_i be the subset of the training set belonging to class w_i and N_i be the cardinality of Z_i . The decision template for class w_i , denoted \mathcal{DT}_i is the mean of the class in the intermediate feature space:

$$\mathcal{DT}_i = \frac{1}{N_i} \sum_{z_j \in Z_i} \mathcal{DP}(z_j). \quad (12)$$

To test a sample \mathbf{x} , we construct $\mathcal{DP}(\mathbf{x})$ and calculate the distance between $\mathcal{DP}(\mathbf{x})$ and each \mathcal{DT}_i using

$$d_E(\mathcal{DP}(\mathbf{x}), \mathcal{DT}_i) = \sum_{j=1}^C \sum_{k=1}^L (d_{kj}(\mathbf{x}) - dt_i(k, j))^2, \quad (13)$$

where $dt_i(k, j)$ is the (k, j) th entry in the decision template \mathcal{DT}_i . The support for class w_i offered by combining the L classifiers, $\text{Conf}_{\mathcal{DT}_i}^i(\mathbf{x})$, is then found by measuring the similarity between the current $\mathcal{DP}(\mathbf{x})$ and \mathcal{DT}_i :

$$\mu_i(\mathbf{x}) = 1 - \frac{1}{L \times C} d_E(\mathcal{DP}(\mathbf{x}), \mathcal{DT}_i). \quad (14)$$

In the landmine detection application, we use the confidence values of the eight algorithms to construct an 8×2 decision templates. We let $d_{i1}(\mathbf{x}) = p_m^i(\mathbf{x})$ and $d_{i2}(\mathbf{x}) = p_c^i(\mathbf{x})$, where p_m^i and p_c^i are the mine and clutter probabilities computed as in Section 4.5. The final mine confidence value is

$$\text{Conf}_{\mathcal{DT}}(\mathbf{x}) = \mu_1(\mathbf{x}) \times (1 - \mu_2(\mathbf{x})). \quad (15)$$

4.7. Rank-based fusion

This approach is based on the voting method proposed by Borda [79]. Each algorithm ranks all the candidate objects in order of their confidences. In particular, each algorithm i maps the confidence value of object x_j , ($y_i(x_j)$), to a rank value r using

$$r_i(x_j) = 1 + \sum_{k \neq j} \chi_{\geq}(y_i(x_j), y_i(x_k)) + \frac{1}{2} \left(\sum_{k \neq j} \chi_{=}(y_i(x_j), y_i(x_k)) \right). \quad (16)$$

In (16), χ_{\geq} is the characteristic function that maps a pair in which the first element is greater than or equal to the second to 1 and all other pairs to 0. Similarly, $\chi_{=}$ maps identical pairs to 1. Thus, each object in the training set will have a rank in the interval $[1, N]$, where N is the size of the training set.

Let $\alpha_i \in \mathbb{R}, i = 1, \dots, L$. The weighted Borda fusion of L algorithms is defined to be weighted sum of the ranks assigned by each algorithm:

$$\text{Conf}_{B_w}(\mathbf{x}) = \frac{1}{N \times L} \sum_{i=1}^L \alpha_i r_i(\mathbf{x}). \quad (17)$$

If $\alpha_i = 1 \forall i$, (17) is called the Borda count and $\text{Conf}_{B_w}(\mathbf{x}) \in [0, 1]$.

Borda fusion has been applied to landmine detection [80], and (in a different way) to handwriting recognition [46], and fusion of social choices (voting, evaluation, etc.) The main advantages of the Borda based fusion is that it makes no assumptions about the underlying distributions of the confidence value assignments. In addition, it maps each of the confidence distribution to a uniform distribution, thus providing a reasonable method for combining decision statistics.

To apply this voting strategy in a supervised learning setting, we rank the training set alarms as shown in (16). Although the

algorithm confidences may depend upon the properties of the training set, the ranking process makes no use of such *a priori* information. Rank values are assigned to test objects using the training set rankings. Thus, if algorithm i assigns confidence x_k to object k , we assign rank $r_i(x_k)$ (the training set rank associated with that algorithm confidence value) to object k .

We have explored weight selection techniques such as Kendall's rank correlation coefficient [81], coefficient of concordance [82], and weights motivated by gambling theory [83]. All of them outperform unweighted Borda fusion. Exhaustive search can be used to assign weights for small collections of algorithms, but is too computationally burdensome for large collections.

Given an assignment of algorithm weights, w , Conf_{B_w} maps each object to its corresponding confidence. Thus, for each vector w , there is a ROC curve. As in the GEOM detector, we seek to maximize the area under the ROC curve. Consider the function $\text{AUC}(w)$, mapping an algorithm weight assignment to the corresponding area under the ROC curve given by Conf_{B_w} . To identify the best weights to use, we perform gradient ascent on $\text{AUC}(w)$ starting with $w_i = 1/L$ for all i . The weights are constrained to sum to 1, but they can be either positive or negative.

4.8. Discrete Choquet integral

The Choquet integral has been investigated for information fusion by many researchers [84–89, 95, 90–93]. This integral defines a family of generally nonlinear aggregation operators on some function of the algorithm confidence values, which we will refer to as a decision statistic. The aggregation operator is defined by the discrete Choquet integral with respect to a non-additive fuzzy measure. As used here, fuzzy measures are real-valued functions defined on sets of algorithms. There are many non-additive measures that can be used with the Choquet integral. The Choquet integral with respect to a specific non-additive measure is a specific aggregation operator such as the *mean*, *median*, *max*, *min*, *trimmed means*, *Ordered Weighted Averaging* operators, and *voting* operators as well as more complex operators. Many of these operators are already used in fusion. The Choquet integral is a mathematical construct that can be used to optimize the aggregation operator for a specific fusion application.

Discrete fuzzy measures and Choquet integrals are defined as follows [94, 86, 8]:

Definition 1. Let $Y = \{y_1, \dots, y_n\}$ be any finite set. A discrete fuzzy measure on Y is a function $\mu: 2^Y \rightarrow [0, 1]$ with the following properties:

- (1) $\mu(\emptyset) = 0$ and $\mu(Y) = 1$.
- (2) Given $A, B \in 2^Y$, if $A \subset B$ then $\mu(A) \leq \mu(B)$ (Monotonicity Property).

Definition 2. Let $f: Y \rightarrow [0, 1]$ and let σ denote a permutation such that $0 \leq f(y_{\sigma(1)}) \leq \dots \leq f(y_{\sigma(n)})$, and let $A_{(i)}$ be given by $A_{(i)} = \{y_{\sigma(i)}, \dots, y_{\sigma(n)}\}$. The Choquet integral of f is:

$$C_\mu(f) = \sum_{i=1}^n \mu(A_{(i)}) (f(y_{(i)}) - f(y_{(i-1)})) = \sum_{i=1}^n f(y_{(i)}) (\mu(A_{(i)}) - \mu(A_{(i+1)})), \quad (18)$$

where we take $f(y_{(0)}) \equiv 0$, $A_{(n+1)} \equiv \emptyset$ and $y_{(i)} \equiv y_{\sigma(i)}$.

In these experiments, algorithm ranks as described in the section on Borda fusion are used as the function f .

Several algorithms have been proposed for learning fuzzy measures [88, 95, 96]. In this paper, we report the results obtained using a learning algorithm that is based on a Bayesian model that

combines logistic regression with sparsity promoting priors [97]. More specifically, this algorithm seeks to maximize the a-posteriori probability of the measure given the data. The posterior probability of the measure is proportional to the product of the likelihood function and the prior probability of the measure. An exponential prior is assumed on the fuzzy measure parameters. Since the probability of a zero parameter is very high with this prior, it is likely to drive measure parameters to zero in the learning process and potentially eliminate unnecessary algorithms from the fusion. The likelihood function is a binomial distribution, and the MAP estimate is computed using a Gibbs sampling algorithm that is designed to maintain the monotonicity constraints of the fuzzy measure [97].

4.9. Context-dependent fusion

The context-dependent fusion (CDF) approach [51] is motivated by the observation that there is no single algorithm that can consistently outperform all others detectors. For instance, in landmine detection, the relative performance of different detectors can vary significantly depending on the mine type, geographical site, soil and weather conditions, and burial depth.

The training part of CDF has two main components: *Context Extraction*, and *Algorithm Fusion*. In *Context Extraction*, the features extracted by the different algorithms are combined, and a clustering algorithm is used to partition the training signatures into groups of similar signatures, or contexts, and learn the relevant features within each context. It is assumed that signatures that have similar response to different algorithms share some common features, and would be assigned to the same cluster. The *Algorithm Fusion* component assigns an aggregation weight to each detector in each context based on its relative performance within the context. To test a new signature using CDF, each detector extracts its set of features and assigns a confidence value. Then, the features are used to identify the best context, and the aggregation weights of this context are used to fuse the individual confidence values.

We should note here that CDF is an alternative approach to data fusion that is local, and that adapts the fusion method to different regions of the feature space. It has been applied to landmine detection [51] using simple linear aggregation. However, any of the fusion methods outlined earlier could be integrated into this approach.

The features extracted by the seven discrimination algorithms from the training alarms are used to partition the feature space into 20 clusters. We use SCAD [98] to do so since it can partition the feature space and learn optimal feature relevance weights for each partition. For each cluster, the seven algorithms and the prescreener are scored separately and a degree of worthiness is assigned to each based on the overlap between the distributions of the mine and clutter confidence values. Algorithms with less overlap are considered more “expert” for the cluster under consideration, and are assigned larger weights. The worthiness of all

eight algorithms are constrained to sum to 1. Let O_k denote the overlap for algorithm k . The degree of worthiness of algorithm k in context i is computed using

$$w_k^i = \frac{\frac{1}{\epsilon + (O_k^i)^2}}{\sum_{j=1}^8 \frac{1}{\epsilon + (O_j^i)^2}}, \quad (19)$$

where ϵ is a small number used to avoid division by zero when the classes are separable. Assuming that alarm x is assigned to context i , its fused confidence value is computed using

$$\text{Conf}_{\text{CDF}}(x) = \sum_{k=1}^8 w_k^i \times y_k, \quad (20)$$

where y_k is the confidence value assigned by algorithm k .

The above fusion methods were selected for evaluation and comparison for the landmine detection application because they have very distinctive properties. For instance, one method (CDF) is local and adapts the detectors’ worthiness to different data subspaces. The other methods are global and assign a degree of worthiness to each detector that is averaged over the entire training data. Also, some fusion methods operate on the detectors’ confidence values of the alarms, others (Borda and fuzzy integral) operate on the ranks of the alarms, and others (CDF) require both confidence values and features used by the classifiers. Another main difference between these fusion methods is the way they are trained. Some methods use straightforward training (e.g. Decision template and Bayes), while others (e.g. fuzzy integral) use more elaborate training algorithms. Moreover, the trainable methods use different optimization criteria. For instance, some try to maximize the area under the ROC, while others minimize the overlap between the distribution of the confidence values in the classes of mines and clutter. These algorithms were developed by various subsets of the authors. Maximal performance for each fusion algorithm was always the goal of the algorithm developer. As is always the case, it is possible that better performance could be found with any of the tested approaches, such as Dempster–Shafer, for example. The characteristics of the different fusion methods are summarized in Table 1.

5. Experimental results

5.1. Dataset statistics

The discrimination algorithms and the different fusion methods were implemented and tested with data collected using the NIITEK vehicle mounted GPR system. The data were collected between November 2002 and July 2006 from four geographically distinct test sites. Sites A, B, and D are temperate climate test facilities with prepared soil and gravel lanes. Site C is an arid climate test facility with prepared soil lanes. The four sites have a total of 17 different lanes with known mine locations. All mines are anti-tank (AT) mines. In all, there are 19 distinct mine types that can be classified

Table 1
Characteristics of the seven fusion methods.

Fusion Alg.	Assumption	Input	Local/global	Considers classifiers correlation	Considers subsets of classifiers	Aggregation weights	Requires training	Optimized criterion
Bayes	Mixture of Gaussian	Conf.	Global	Yes	No	N/A	Yes	Log likelihood
Mahalanobis distance	Gaussian distribution	Conf.	Global	Yes	No	N/A	Yes	Average FAR for PD ∈ [92%, 96%]
Dempster–Shafer	N/A	Conf.	Global	No	No	Positive	Yes	N/A
Decision template	N/A	Conf.	Global	No	No	N/A	Yes	Distance to decision template
Borda	N/A	Rank	Global	No	No	Positive/negative	Yes	Area under ROC
Fuzzy integral	N/A	Rank	Global	No	Yes	Positive	Yes	Posterior prob. of the measures
Context-dependent	N/A	Conf. + Feat	Local	No	No	Positive	Yes	Class dist. overlap

Table 2
Statistics of the dataset.

	Site A	Site B	Site C	Site D	Total
No. collections	3	6	2	1	12
No. mine types	9	15	9	5	19
No. mine alarms	183	821	62	494	1560
No. clutter encounters	0	15	0	196	211
No. clutter alarms post prescreener	0	4	0	46	50
Area (m ²)	14,813	15,631	4054	7310	41,808

Table 3
Number of metal and plastic cased mines and mine simulants and their burial depths.

	Depth								Total
	–1 in.	0 in.	1 in.	2 in.	3 in.	4 in.	5 in.	6 in.	
Metal	12	37	124	68	151	34	119	77	777
Low-metal	6	92	90	204	122	134	47	76	616
Simulants	48	0	20	47	23	29	0	0	167
Total	66	129	234	319	296	197	166	153	1560

into three categories: anti-tank metal (ATM), anti-tank with low metal content (ATLM), and simulated mines (SIM). The targets were buried up to 6 in. deep. Multiple data collections were performed at each site at different dates, covering a ground area of 41,807.57 m², resulting in a large and diverse collection of mine and false alarm signatures. False alarms arise as a result of radar signals that present a mine-like character. Such signals are generally said to be a result of clutter. In this experiment, clutter arises from two different processes. One type of clutter is emplaced and surveyed in an effort to test the robustness of the algorithms. Other clutter result from human activity unrelated to the data collection or as a result of natural processes. We refer to this second kind of clutter as non-emplaced. Non-emplaced clutter includes objects discarded or lost by humans, soil inconsistencies and voids, stones, roots and other vegetation, as well as remnants of animal activity.

The statistics of the data are shown in Table 2. The data collected from Sites B and D have emplaced buried clutter. Although the lanes at Sites A and C are prepared, they still contain non-emplaced clutter objects. Both metal and non-metal non-emplaced clutter objects such as ploughshares, shell casings, and large rocks have been excavated from these sites. The emplaced clutter objects include steel scraps, bolts, soft-drink cans, concrete blocks, plastic bottles, wood blocks, and rocks. In all, there are 12 collections having 19 distinct mine types. Many of these mine types are present at several sites. The prescreener detected 1560 of the 1593 mines encountered in the data, yielding a 97.9% probability of detection. It rejected 161 of 211 emplaced clutter objects encountered, and yielded a total of 3435 false alarms associated with non-emplaced clutter objects. The number, type, and burial depth of the mines are given in Table 3. As it can be seen, the mines buried at 1 inch through 6 inches occupy 87.5% of the total targets encountered vs. 12.5% surface-laid or flush-buried mines.

5.2. Implementation issues

Each of the seven detection algorithms (EHD, HMM, Spect, Geom, TFCM, GFIT, and GMRF) and the seven fusion methods (context-dependent, Bayes, decision template, Dempster–Shafer, Mahalanobis distance, fuzzy integral, and Borda count) were implemented for use with the Testing/training Unified Framework (TUF) system. This system supports creation of supervised learning algorithms that perform discrimination between targets and non-targets in data collected at a variety of different regions (mine lanes) in a variety of different sites. The framework employs algo-

rithms implemented in Matlab using a control flow that incorporates a user-programmed prescreener (NUKEv6) that processes raw data files into alarms with associated Universal Transverse Mercator (UTM) coordinates and confidence values. The alarms are then processed by extracting signatures. These signatures are passed to a user-specified feature extractor. The features resulting from the feature extractor are presented along with the alarms to a discrimination algorithm, which produces a confidence for each alarm. The system performs *n*-way cross validation testing using either lane-based cross validation (in which each mine lane is in turn treated as a test set with the rest of the lanes used for training) or site-based cross validation (in which each data collection site is treated in turn as a test set). The EHD, Geom, TFCM, GFIT, and GMRF detection algorithms are trained in this cross validation manner. The HMM was based on a model trained using a different radar system and the Spect employs a single static mine model and is not trained. For the fusion, all algorithms are trained and tested using the same cross validation scheme.

5.3. Evaluation method

To provide an objective and consistent evaluation of all algorithms, we use the TUF system with lane-based cross validation. The results of this process are scored using the MIne Detection Assessment and Scoring (MIDAS) system developed by the Institute for Defense Analysis [99]. The scoring is performed in terms of probability of detection (PD) vs. false alarm rate (FAR). Confidence values are thresholded at different levels to produce Receiver Operating Characteristic (ROC) curve. For a given threshold, a mine is detected if there is an alarm within 0.25 m from the edge of the mine with confidence value above the threshold. Given a threshold, the PD is defined to be the number of mines detected divided by the number of mines. The FAR is defined as the number of false alarms per square meter.

5.4. Results and analysis

5.4.1. Individual detection algorithms

First, we compare the performance of the individual detectors and justify the need to fuse their results to improve the overall performance of the system. Fig. 3 displays the ROC's obtained by applying the seven detection algorithms and the prescreener to the entire data collection. As it can be seen, the EHD detector has the best overall performance. However, this does not necessarily mean that the EHD is *consistently* the best algorithm. For instance, Fig. 4a displays the results averaged over site A of the collection only. For this subset, the EHD is the best algorithm and the HMM

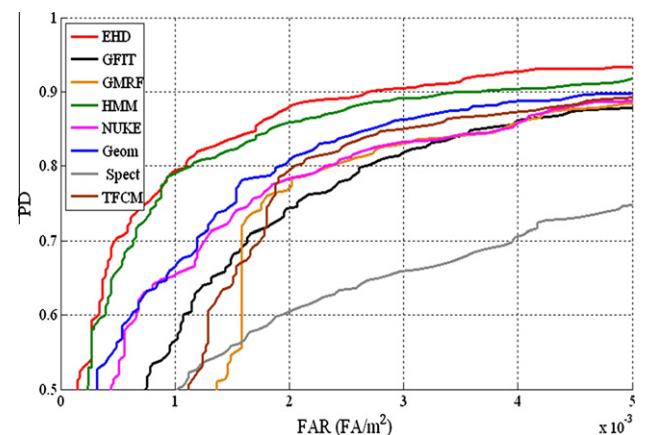


Fig. 3. Performance of the eight different detectors on the entire data collection.

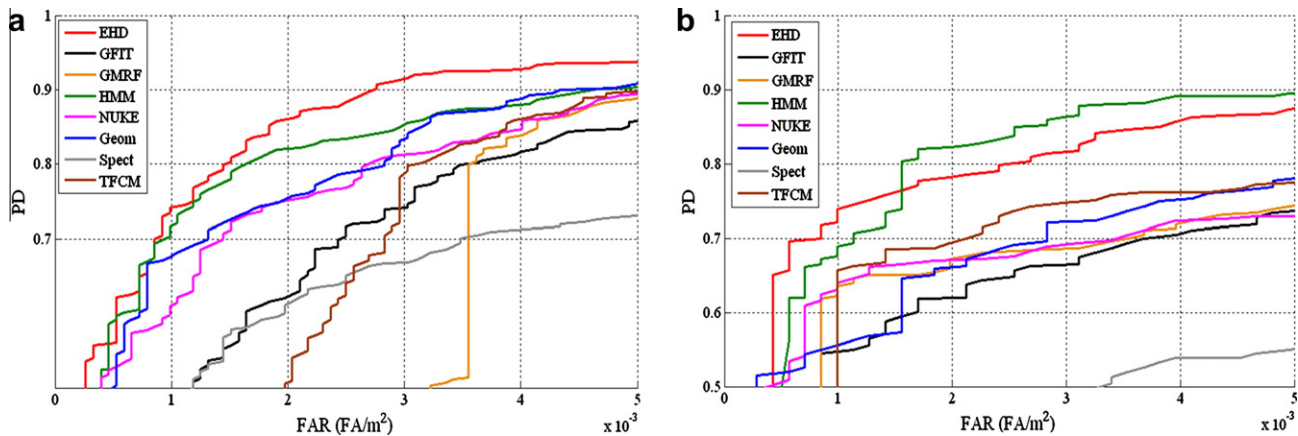


Fig. 4. Performance of the eight detectors on: (a) Site A only; and (b) Site B only.

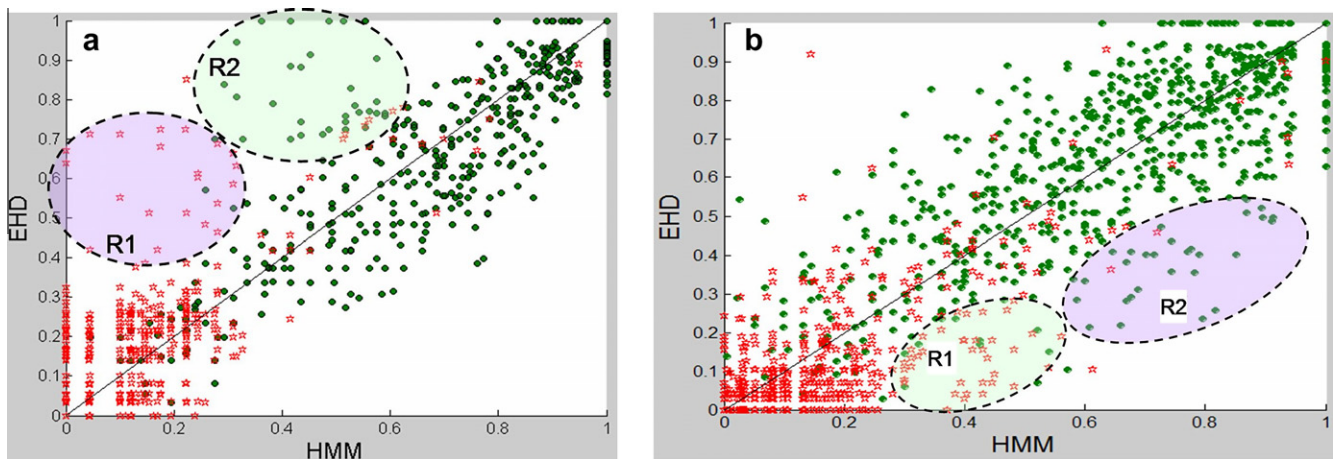


Fig. 5. Comparison of the EHD and HMM outputs for several mine (green dots) and clutter (red stars) signatures extracted from: (a) a subset of Site A; and (b) a subset of Site B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is the second best one. However, in Fig. 4b, which displays the results averaged over site B only, the HMM is the best algorithm and EHD is the second best one.

Thus, there is no single algorithm that can consistently outperform all others detectors. In fact, the relative performance of different detectors can vary depending on the geographical site and soil and weather conditions. Moreover, even within the same site, the relative performance of the different algorithms can vary significantly depending on the mine type, burial depth, and other unknown factors. To illustrate this, we compare the output of the HMM and EHD detectors for a small subset of alarms extracted from the same site in Fig. 5. For instance, the highlighted region (R1) in Fig. 5a includes mainly clutter signatures where the HMM algorithm outperforms the EHD (lower HMM confidence values). On the other hand, for the same subset, region (R2) includes mainly mine signatures where the EHD detector outperforms the HMM (higher EHD confidence). Fig. 5b highlights two other regions for another geographical site.

5.4.2. Fusion results

Our objective is to evaluate a set of fusion methods to combine the output of several landmine discrimination algorithms to determine their suitability for use in an automated detection system in a variety of locations and under different environments. In addition to the performance of these fusion methods, we are also interested in their scalability with respect to the number of discrimination

algorithms. Thus, we compare these methods when 4, 6, and 8 discrimination algorithms are considered.

Fig. 6 displays the results of the seven fusion algorithms when only four discrimination algorithms (EHD, HMM, Spect, and NUKE) are combined.

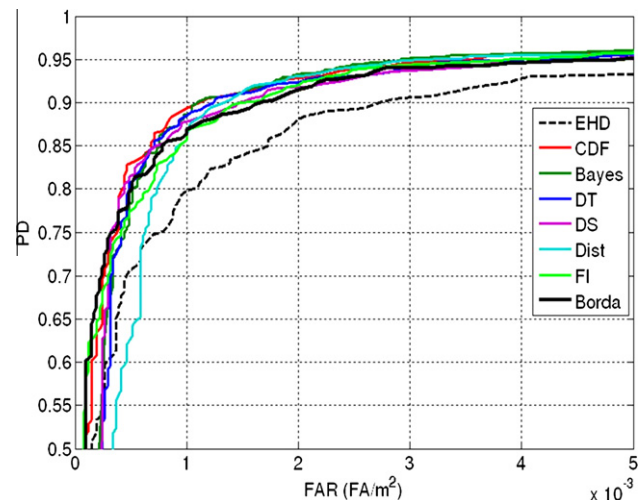


Fig. 6. Comparison of seven fusion methods when four discrimination algorithms (EHD, HMM, Spect, and NUKE) are combined.

are fused. We also include the ROC of the EHD (best overall discrimination algorithm) as a reference. As it can be seen, the ROC's of all fusion methods are clustered together, and thus all methods have comparable performances. All fusion methods improve the PD results over the best discrimination algorithm by an average of 10% for FAR around 0.0007. At low PD (<80%), the Mahalanobis distance based fusion results are not as good as the other methods. This is due mainly to the fact that one single Gaussian component may not be sufficient to model the distribution of the confidence values of the individual discriminators in the four-dimensional confidence space. The Bayes-based method, which is similar to the distance based, does not exhibit this behavior because multiple Gaussian components (M was estimated to be 4) were used to model the distribution of each class. It is also interesting to note that the distance based fusion outperforms Bayes at higher PD. This is because the former method is optimized to minimize the average FAR for $PD \in [92\%, 96\%]$.

Fig. 7 displays the results of the seven fusion algorithms when only six discrimination algorithms (EHD, HMM, Spect, NUKE, Geom, and TFCM) are fused. First, we notice the addition of two discrimination algorithms did not improve the results of any of the fusion methods. Two possible reasons may explain this behavior. First, the added discrimination algorithms (TFCM and Geom) are based on edge, texture, and geometric features that are already used (in a different way) by the other discrimination algorithms. Second, it is possible that for the data collection that was used, it is not possible to improve the results further.

Comparing the results in Fig. 7 to those in Fig. 6, we observe that for some fusion methods, the performance has degraded. In particular, the performance of the Dempster–Shafer (DS) and the decision template (DT) methods have dropped significantly at low PD (<80%) and have become even worse than the EHD discriminator. Investigation of this problem has revealed that these two fusion methods generate confidence values that have a distribution close to binary. This behavior is due to the way the basic belief functions are aggregated (refer to Eq. (9)). In particular, adding more algorithms will require more multiplications. For the DT method, the dimension of the decision template matrix increases, and this may drive the distances in (13) to a bimodal distribution. Due to these nearly binary distributions, weak mines will be assigned confidence values close to zero, and this would explain the lower PD at low FAR. Also, strong false alarms will be assigned confidence values close to 1, and this would explain the relatively lower PD at higher FAR.

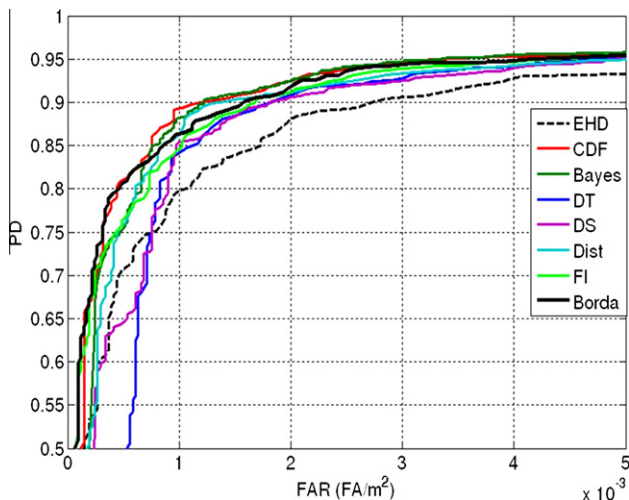


Fig. 7. Comparison of seven fusion methods when six discrimination algorithms (EHD, HMM, Spect, NUKE, Geom, and TFCM) are combined.

Fig. 8 compares the results of the seven fusion algorithms when eight discrimination algorithms (EHD, HMM, Spect, NUKE, Geom, TFCM, GFIT, and GMRF) are fused. First, we note that the performance of the DT and DS degraded further as the confidence values become closer to binary. Second, the performance of all other fusion methods (except CDF) have degraded compared to the fusion of four algorithms only. This may be due to the fact that the four added algorithms have lower performances (refer to Fig. 3), and when all eight algorithms are fused globally, the added algorithms have a negative impact. Third, we note that the dependency assumption does not seem to be an issue. In fact, the two best fusion methods (CDF and Borda) assume that the eight discrimination algorithms are independent.

The Borda count fusion is the second best method, and does not seem to be affected by the addition of discrimination algorithms. This is due to the fact that this method allows for negative aggregation weights as long as they improve the area under the ROC. Thus, as we add more discrimination algorithms (with worse overall performance), this method will assign negative (or zero) weights to these algorithms.

The CDF has the best overall performance. Moreover, the addition of discrimination algorithms did not degrade its performance. In fact, for certain FAR values, its performance has improved. This is due to the fact that this method is local and strives to take advantage of the different detectors in different contexts. For any cluster (or context) the detectors are ranked based on the overlap between the mine and clutter confidence distribution. This ranking can ignore (by assigning low aggregation weights) many of the discrimination algorithms. It could also assign a significant weight to discrimination algorithms that are good for the given context, but globally, are not as good as other algorithms. We have observed that on average, this fusion assigns significant aggregation weights to 3–5 discrimination algorithms. These algorithms differ from one cluster to another.

Finally, we should note the fuzzy integral approach is trained using a learning algorithm that combines logistic regression with sparsity promoting priors. Thus, it is designed to ignore individual discrimination algorithms that do not improve the results. However, the results do not seem to support this. This may be due to the fact that the number of parameters increases exponentially as we increase the number of algorithms. Thus, the search for the optimal parameters becomes more complex and may lead to sub-optimal solutions.

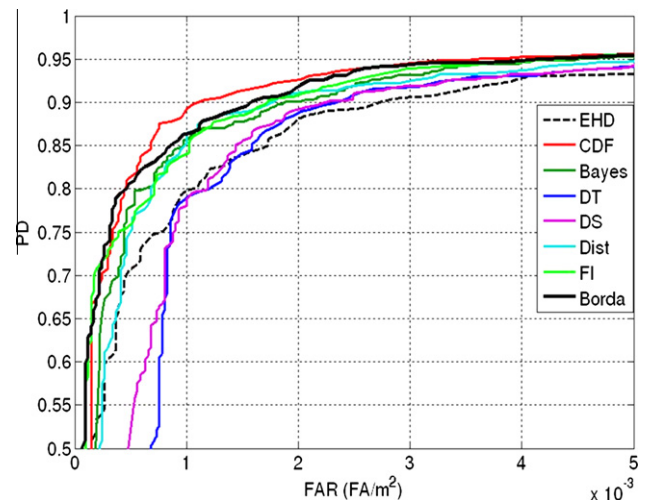


Fig. 8. Comparison of seven fusion methods when eight discrimination algorithms (EHD, HMM, Spect, NUKE, Geom, TFCM, GFIT, and GMRF) are combined.

6. Conclusions

We have presented results of an evaluation of several fusion methods to combine the output of several anti-tank landmine discrimination algorithms. Our objective was to determine the suitability of these methods for use in an automated detection system in a variety of locations and under different environments. Our extensive research and testing in this application has revealed that algorithm performances for buried anti-tank landmine detection are strongly dependent upon a variety of factors that are not well understood. It is typically the case that one algorithm may perform well in one setting and not so well in another. Thus, in order to achieve a reliable and robust detection system, several distinct detection algorithms need to be developed and fused. Therefore, in addition to the performance of the different fusion methods, we are also interested in their scalability with respect to the number of discrimination algorithms. In particular, their ability to take advantage of discrimination algorithms that perform well for only a small subset of the data without being affected by their weakness. To investigate this, we have compared the seven fusion methods when 4, 6, and 8 discrimination algorithms are considered.

Our experimental results show that although the fusion algorithms were all quite similar when a small number of algorithms were fused, the performance was more varied as the number of algorithms increased. Context-dependent fusion appears to be an excellent approach that should be investigated in more detail in future work. Aggregation operators that are allowed to use negative weights appear to perform better than those that do not. Sparsity promoting priors do not necessarily lead to better performance as the number of algorithms increases. The tradeoff between promoting sparsity and computational complexity is difficult to control. Fusion algorithms that tend to binarize confidence values as the number of inputs increases also degraded as a function of the number of algorithms fused. The assumption that the individual detectors are statistically independent does not seem to be a significant factor in affecting the performance of the fusion methods. However, this may be an important issue should the need to reduce the overall computational requirements of the system arises. Future work will look at integrating the Bayes, Dempster–Shafer, and the Choquet fusion methods within the context-based fusion concept.

Acknowledgment

The authors thank R. Harmon, R. Weaver, P. Howard, and T. Donzelli for their support of this work, E. Rosen and L. Ayers of IDA for useful software and insight. We also thank L. Carin, L. Collins and P. Torrione of Duke University and NIITEK, Inc., for their insights, cooperation, discrimination algorithms, and data. This work was supported in part by NSF Awards No. CBET-0730802 and CBET-0730484, ONR Award Number N00014-05-10788, ARO and ARL Cooperative Agreement Number DAAD19-02-2-0012 and Grant Number DAAB15-02-D-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Office of Naval Research, Army Research Laboratory, or the US Government.

References

- [1] Hidden Killers, The Global Landmine Crisis, United States Department of State Report, Publication No. 10575, September 1998.
- [2] J.A. MacDonald, Alternatives for Landmine Detection, RAND Corporation, 2003.
- [3] J.N. Wilson, P. Gader, W. Lee, H. Frigui, K.C. Ho, A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine

- detection and discrimination, *IEEE Transactions on Geoscience and Remote Sensing* 45 (2007) 2560–2572.
- [4] S.L. Tantom, Y. Wei, V.S. Munshi, L.M. Collins, A comparison of algorithms for landmine detection and discrimination using ground penetrating radar, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets*, 2002, pp. 728–735.
- [5] P. Gader, B. Nelson, H. Frigui, G. Vailllette, J. Keller, Fuzzy logic detection of landmines with ground penetrating radar, *Signal Processing* 80 (2000) 1069–1084 (special issue on fuzzy logic in signal processing).
- [6] P. Gader, M. Mystkowski, Y. Zhao, Landmine detection with ground penetrating radar using hidden markov models, *IEEE Transactions on Geoscience and Remote Sensing* 39 (2001) 1231–1244.
- [7] H. Frigui, P. Gader, K. Satyanarayana, Landmine detection with ground penetrating radar using fuzzy k-nearest neighbors, in: *Proceedings of the IEEE Conference on Fuzzy Systems*, Budapest, Hungary, 2004, pp. 1745–1749.
- [8] P. Gader, L. Wen-Hsiung, A. Mendez-Vazquez, Continuous Choquet integrals with respect to random sets with applications to landmine detection, in: *IEEE International Conference on Fuzzy Systems*, 2004, pp. 523–528.
- [9] P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, A. Zare, Multi-sensor and algorithm fusion with the Choquet integral: applications to landmine detection, in: *Geoscience and Remote Sensing Symposium*, vol. 1, 2004, pp. 1605–1608.
- [10] P. Torrione, L. Collins, Application of texture feature classification methods to landmine and clutter discrimination in off-road GPR data, in: *Geoscience and Remote Sensing Symposium*, vol. 1, 2004, pp. 1621–1624.
- [11] S. Sheedvash, M. Azimi-Sadjadi, Structural adaptation in neural networks with applications to land mine detection, in: *IEEE International Conference on Neural Networks*, 1997, pp. 1443–1447.
- [12] Q.L.J. Zhang, B. Nath, Landmine feature extraction and classification of GPR data based on SVM method, in: *International Symposium on Neural Networks*, 2004, pp. 636–641.
- [13] X. Miao, M. Azimi-Sadjadi, B. Tian, A. Dubey, N. Witherspoon, Detection of mines and minelike targets using principal component and neural methods, in: *IEEE International Conference on Neural Networks*, 1998, pp. 454–463.
- [14] C. Yang, Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset, *IEEE Transactions on Neural Networks* 16 (3) (2005) 743–753.
- [15] O. Lohlein, M. Fritzsche, Classification of GPR data for mine detection based on hidden markov models, in: *EUREL Conference on the Detection of Abandoned Landmines*, 1998, pp. 96–100.
- [16] T.R. Witten, Present state of the art in ground-penetrating radars for mine detection, in: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, 1998, pp. 576–586.
- [17] P.D. Gader, H. Frigui, B. Nelson, G. Vailllette, J.M. Keller, New results in fuzzy set based detection of landmines with GPR, in: *Detection and Remediation Technologies for Mines and Minelike Targets IV*, 1999, pp. 1075–1084.
- [18] H.T. Kaskett, J.T. Broach, Automatic mine detection algorithm using ground penetrating radar signatures, in: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets*, 1999, pp. 942–952.
- [19] E. Rosen, Investigation into the sources of persistent ground-penetrating radar false alarms: data collection, excavation, and analysis, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets VIII*, 2003, pp. 185–190.
- [20] P.A. Torrione, C.S. Throckmorton, L.M. Collins, Performance of an adaptive feature-based processor for a wideband ground penetrating radar system, *IEEE Transactions on Aerospace and Electronic Systems* 42 (2) (2006) 644–657.
- [21] P. Gader, W.H. Lee, J.N. Wilson, Detecting landmines with ground penetrating radar using feature-based rules, order statistics, and adaptive whitening, *IEEE Transactions on Geoscience and Remote Sensing* 42 (11) (2004) 2522–2534.
- [22] K.J. Hintz, Snr improvements in NIITEK ground penetrating radar, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IX*, 2004, pp. 399–408.
- [23] D. Carevic, Clutter reduction and target detection in ground penetrating radar data using wavelets, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV*, 1999, pp. 973–978.
- [24] D. Carevic, Kalman filter-based approach to target detection and target-background separation in ground-penetrating radar data, in: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV*, 1999, pp. 1284–1288.
- [25] A. Gunatilaka, B.A. Baertlein, Subspace decomposition technique to improve GPR imaging of anti-personnel mines, in: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets V*, 2000, pp. 1008–1018.
- [26] H. Brunzell, Detection of shallowly buried objects using impulse radar, *IEEE Transactions on Geoscience and Remote Sensing* 37 (1999) 875–886.
- [27] S. Yu, R.K. Mehra, T.R. Witten, Automatic mine detection based on ground penetrating radar, in: *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV*, 1999, pp. 961–972.
- [28] H. Frigui, K.C. Ho, P. Gader, Real-time land mine detection with ground penetrating radar using discriminative and adaptive hidden markov models, *EURASIP Journal on Applied Signal Processing* 12 (2005) 1867–1885.
- [29] H. Frigui, P.D. Gader, Detection and discrimination of land mines based on edge histogram descriptors and fuzzy k-nearest neighbors, in: *Proceedings of the IEEE International Conference on Fuzzy Systems*, Vancouver, BC, Canada, 2006, pp. 1494–1499.

- [30] K.C. Ho, L. Carin, P.D. Gader, J.N. Wilson, An investigation of using the spectral characteristics from ground penetrating radar for landmine/clutter discrimination, *IEEE Geoscience and Remote Sensing Letters* 46 (4) (2008) 1177–1191.
- [31] P.D. Gader, W.-H. Lee, J.N. Wilson, Detecting landmines with ground penetrating radar using feature-based rules order statistics, and adaptive whitening, *IEEE Transactions on Geoscience and Remote Sensing* 42 (11) (2004) 2522–2534.
- [32] W.-H. Lee, P.D. Gader, J.N. Wilson, Optimizing the area under a receiver operating characteristic curve with application to landmine detection, *IEEE Transactions on Geoscience and Remote Sensing* 45 (2) (2007) 389–397.
- [33] P. Torrione, L.M. Collins, Texture features for antitank landmine detection using ground penetrating radar, *IEEE Transactions on Geoscience and Remote Sensing* 45 (7) (2007) 2374–2382.
- [34] M.-H. Horng, Texture feature coding method for texture classification, *Optical Engineering* 42 (1) (2003) 228–238.
- [35] P.A. Torrione, L. Collins, Application of Markov random fields to landmine detection in ground penetrating radar data, in: *Proceedings of the SPIE Conference on Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII*, vol. 6953, 2008, pp. 69531B–695312.
- [36] P. Torrione, personal communication.
- [37] L. Rastgrin, R. Erensterin, *Method of Collective Recognition*, Energoizdat, Moscow, Russian, 1981 (in Russian).
- [38] R. Jacobs, Methods for combining experts probability assessments, *Neural Computation* 7 (5) (1995) 867–888.
- [39] C. Ji, S. Ma, Combined weak classifiers, in: M. Mozer, M. Jordan, T.E. Petsche (Eds.), *Advances in Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, 1997, pp. 494–500.
- [40] T. Ho, J. Hull, S. Srihari, Decision combination in multiple classifier systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (1994) 66–75.
- [41] P. Munro, B. Parmanto, Combining neural network regression estimates with regularized linear weights, in: M. Mozer, M. Jordan, T.E. Petsche (Eds.), *Advances in Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, 1997, pp. 592–598.
- [42] S. Hashem, Optimal linear combinations of neural networks, *Neural Networks* 10 (4) (1997) 599–614.
- [43] L. Lam, C. Suen, Optimal combination of pattern classifiers, *Pattern Recognition Letters* 16 (1995) 945–954.
- [44] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (3) (1998) 226–239.
- [45] H. Tahani, J.M. Keller, Information fusion in computer vision using the fuzzy integral, *IEEE Transactions on Systems Man and Cybernetics* 20 (3) (1990) 733–741.
- [46] P.D. Gader, M.A. Mohamed, J.M. Keller, Fusion of handwritten word classifiers, *Pattern Recognition Letters* 17 (6) (1996) 577–584.
- [47] S. Le Hegarat-Masclé, I. Bloch, D. Vidal-Madjar, Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover, *Pattern Recognition* 31 (11) (1998) 1811–1823.
- [48] N. Milisavljevic, I. Bloch, Sensor fusion in anti-personnel mine detection using a two-level belief function model, *IEEE SMC, Part C: Applications and Reviews* 33 (2003) 269–283.
- [49] E. Mandler, J. Schurmann, Combining the classification results of independent classifiers based on the Dempster–Shafer theory of evidence, *Pattern Recognition and Artificial Intelligence* (1988) 381–393.
- [50] L. Kuncheva, Switching between selection and fusion in combining classifiers: an experiment, *IEEE Transactions on Systems, Man, and Cybernetics – Part B* 32 (2) (2002) 146–156.
- [51] H. Frigui, L. Zhang, P. Gader, D. Ho, Context-dependent fusion for landmine detection with ground penetrating radar, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IX*, 2007, p. 655321.
- [52] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, A. Gelzinis, Soft combination of neural classifiers: a comparative study, *Pattern Recognition Letters* 20 (1999) 429–444.
- [53] L. Kuncheva, Change-glasses approach in pattern recognition, *Pattern Recognition Letters* 14 (1993) 619–623.
- [54] K. Woods, W. Kegelmeyer, K. Bowyer, Combination of multiple classifiers using local accuracy estimates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (4) (1997) 405–410.
- [55] L. Klein, *Sensor and Data Fusion Concepts and Applications*, SPIE, 1993.
- [56] H. Wu, Ph.D. Thesis *Sensor Data Fusion for Context-Aware Computing Using Dempster–Shafer Theory*, 2003.
- [57] S. Challa, D. Koks, Bayesian and Dempster–Shafer fusion, *Sadhana* 29 (2) (2004) 145–174.
- [58] D.M. Buede, P. Girardi, Information fusion in computer vision using the fuzzy integral, *IEEE Transactions on Systems, Man and Cybernetics – Part A* 27 (5) (1999) 569–577.
- [59] D. Fasbender, J. Radoux, P. Bogaert, Bayesian data fusion for adaptable image pansharping, *IEEE Transactions on Geoscience and Remote Sensing* 46 (6) (2008) 1847–1857.
- [60] F. Cremer, E. Breejen, K. Schutte, Sensor data fusion for anti-personnel landmine detection, in: *Proceedings of the International Conference on Data Fusion (EuroFusion98)*, 1998, pp. 55–60.
- [61] E. Breejen, K. Schutte, F. Cremer, Sensor fusion for anti personnel landmine detection: a case study, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV*, 1999, pp. 1235–1245.
- [62] H. Frigui, R. Krishnapuram, Clustering by competitive agglomeration, *Pattern Recognition* 30 (7) (1997) 1223–1232.
- [63] K.C. Ho, P.D. Gader, H. Frigui, J.N. Wilson, Confidence level fusion of edge histogram descriptor, hidden markov model, spectral correlation feature, and nukev6, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets XII*, 2007, pp. 6553–20.
- [64] M.K. Steven, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, 1998.
- [65] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *The Annals of Statistics* (28) (1967) 325–339.
- [66] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1996.
- [67] L. Xu, A. Krzyzak, C.Y. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, *IEEE Transactions on Systems, Man and Cybernetics* 22 (3) (1992) 418–435.
- [68] M. Beynon, D. Cosker, A.D. Marshall, Methods of combining multiple classifiers and their applications to handwriting recognition, *Expert Systems with Applications* 20 (4) (2001) 357–367.
- [69] Y.A. Aslandogan, C.T. Yu, Evaluating strategies and systems for content based indexing of person images on the web, in: *Proceedings of the ACM International Multimedia Conference and Exhibition*, 2000, pp. 313–321.
- [70] K. Sentz, Combination of Evidence in Dempster–Shafer Theory, Technical Report, Sand 2002-0835.
- [71] R. Yager, On the Dempster–Shafer framework and new combination rules, *Information Sciences* 41 (1987) 93–137.
- [72] L.A. Zadeh, A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination, *The AI Magazine* 7 (1987) 85–90.
- [73] C. Lee, A comparison of two evidential reasoning schemes, *Artificial Intelligence* 35 (1) (1988) 127–134.
- [74] P.L. Bolger, Shafer–Dempster reasoning with applications to multisensor target identification systems, *IEEE Transactions on Systems, Man and Cybernetics* 22 (6) (1987) 968–977.
- [75] L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, Decision templates for multiple classifier fusion: an experimental comparison, *Pattern Recognition* 34 (2) (2001) 299–314.
- [76] C. Dietrich, G. Palm, F. Schwenker, Decision templates for the classification of time series, *Information Fusion* 4 (2) (2003) 101–109.
- [77] F.R.J. Kittler, M. Ballette, J. Czyz, L. Vandendorpe, Decision level fusion of intramodal personal identity verification experts, in: *International Workshop on Multiple Classifier Systems*, 2002, pp. 314–324.
- [78] G. Giacinto, F. Roli, L. Didaci, Fusion of multiple classifiers for intrusion detection in computer networks, *Pattern Recognition Letters* 24 (12) (2003) 1795–1803.
- [79] J.-C. de Borda, *Memoire sur les elections au scrutin*, Histoire de l'Académie Royale des Sciences, Paris, 1781.
- [80] J. Wilson, P. Gader, Use of the Borda count for landmine discriminator fusion, in: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IX*, 2007, p. 655322.
- [81] M.G. Kendall, A new measure of rank correlation, *Biometrika* 30 (1/2) (1938) 81–93.
- [82] M.G. Kendall, B.B. Smith, The problem of m rankings, *Annals of Mathematical Statistics* 10 (3) (1939) 275–287.
- [83] T. Cover, J. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
- [84] S. Auephanwirayakul, J. Keller, P.D. Gader, Generalized Choquet fuzzy integral fusion, *Information Fusion* 3 (1) (2002) 69–85.
- [85] J.-H. Chiang, P. Gader, Hybrid fuzzy-neural systems in handwritten word recognition, *IEEE Transactions on Fuzzy Systems* 5 (4) (1997) 497–510.
- [86] P.D. Gader, B. Nelson, A. Hocaoglu, S. Auephanwirayakul, M. Khabou, Neural versus heuristic development of Choquet fuzzy integral fusion algorithms for land mine detection, in: H. Bunke, A. Kandel (Eds.), *Neuro-fuzzy Pattern Recognition*, World Scientific Publ. Co., 2000, pp. 205–226.
- [87] M. Grabisch, Fuzzy integral for classification and feature extraction, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), *Fuzzy Measures and Integrals, Theory and Applications*, Physica Verlag, 2000, pp. 348–374.
- [88] M. Grabisch, A new algorithm for identifying fuzzy measures and its application to pattern recognition, in: *Fourth IEEE International Conference on Fuzzy Systems*, Yokohama, Japan, 1995, pp. 145–150.
- [89] M. Grabisch, J. Nicolas, Classification by fuzzy integral: performance and tests, *Fuzzy Sets and Systems* 65 (2–3) (1994) 255–271.
- [90] K. Xu, Z. Wang, P.-A. Heng, K.-S. Leung, Classification by nonlinear integral projections, *IEEE Transactions on Fuzzy Systems* 11 (2) (2003) 187–2001.
- [91] A. Temko, D. Macho, C. Nadeu, Fuzzy integral based information fusion for classification of highly confusable non-speech sounds, *Pattern Recognition* 41 (5) (2008) 1814–1823.
- [92] H. Némour, Y. Chibani, Neural network combination by fuzzy integral for robust change detection in remotely sensed imagery, *EURASIP Journal on Advances in Signal Processing* 2005 (1) (2005) 2187–2195.
- [93] H. Frigui, Interactive image retrieval using fuzzy sets, *Pattern Recognition Letters* 22 (9) (2001) 1021–1031.

- [94] M. Grabisch, Modelling data by the Choquet integral, in: V. Torra (Ed.), *Information Fusion in Data Mining*, Physica Verlag, Heidelberg, 2003, pp. 135–148.
- [95] M. Grabisch, H. Nguyen, E. Walker, *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*, Kluwer Academic Publishers, Dordrecht, 1995.
- [96] A. Mendez-Vazquez, P. Gader, J.M. Keller, K. Chamberlin, Minimum classification error training for Choquet integrals with applications to landmine detection, *IEEE Transactions on Fuzzy Systems* 16 (1) (2008) 225–238.
- [97] A. Mendez-Vasquez, Ph.D. Dissertaion, *Information Fusion and Sparsity Promotion Using Choquet Integrals*.
- [98] H. Frigui, O. Nasraoui, Unsupervised learning of prototypes and attribute weights, *Pattern Recognition Journal* 37 (2004) 567–581.
- [99] L. Ayers, E. Rosen, *MIDAS: Mine Detection Assessment and Scoring User's Manual V1.1*, Institute for Defense Analysis, Technical Report, 2004.

Detection and Discrimination of Land mines based on Edge Histogram Descriptors and Fuzzy K-Nearest Neighbors

Hichem Frigui and Paul Gader

Abstract—This paper describes an algorithm for land mine detection using sensor data generated by a ground penetrating radar (GPR) system. The GPR produces a 3-D array of intensity values, representing a volume below the surface of the ground. First, a computationally inexpensive pre-screening algorithm is used to focus attention and identify regions with subsurface anomalies. The identified regions of interest are then processed by a feature extraction algorithm to capture their salient features. We use translation invariant features that are based on the local edge distribution of the 3-D GPR signatures. Finally, a fuzzy K-nearest neighbor rule is used to assign a confidence value to distinguish true detections from false alarms. The proposed algorithm is applied to data acquired from three outdoor test sites at different geographic locations.

I. INTRODUCTION

Detection and removal of landmines is a serious problem affecting civilians and soldiers worldwide. It is estimated that more than 100 million landmines are buried in more than 80 countries around the world, and that 26,000 people, mostly civilians, a year are either killed or maimed by a landmine [1], [2]. The detection problem is compounded by the large variety of landmine types, differing soil conditions, temperature and weather conditions, and varying terrain, to name a few. Traditional fielded approaches use metal detectors. Unfortunately, many modern landmines are made of plastic and contain little or no metal.

A variety of sensors have been proposed or are under investigation for landmine detection. It is necessary to have a very high detection rate with a low false alarm rate. The research problem for sensor data analysis is to determine how well signatures of landmines can be characterized and distinguished from other objects under the ground using returns from one or more sensors. Ground Penetrating Radar (GPR) offers the promise of detecting landmines with little or no metal content. Unfortunately, landmine detection via GPR has been a difficult problem[3], [4]. Although systems can achieve high detection rates, they have done so at the expense of high false alarm rates.

Automated detection algorithms can generally be broken down into four phases: pre-processing, feature extraction, confidence assignment, and decision-making. Pre-processing algorithms perform tasks such as normalization of the data, corrections for variations in height and speed, removal of stationary effects due to the system response, etc. Methods

that have been used to perform this task include wavelets and Kalman filters[5], [6], subspace methods and matching to polynomials [7], and subtracting optimally shifted and scaled reference vectors [8]. Feature extraction algorithms reduce the pre-processed raw data to form a lower-dimensional, salient set of measures that represent the data. Principal component (PC) transforms are a common tool to achieve this task [9], [10]. Confidence assignment algorithms can use methods such as hidden Markov Models [11], [12], fuzzy logic [13], rules and order statistics [14], neural networks, or nearest neighbor classifiers to assign a confidence that a mine is present at a point. Decision-making algorithms often post-process the data to remove spurious responses and use a set of confidence values produced by the confidence assignment algorithm to make a final mine/no-mine decision.

In this paper, we propose a feature-based algorithm for land mine detection in GPR data that uses edge histogram descriptors (EHD) for feature extraction and a fuzzy K-Nearest Neighbors (K-NN) based rule for confidence assignment. First, an adaptive least mean squares (LMS) pre-screener is used to focus attention and identify regions with subsurface anomalies. The identified candidates are processed further by the feature-based discrimination algorithm to attempt to separate mine targets from naturally occurring clutter. A set of alarms with known ground truth is used to train the decision making process. These alarms are clustered to identify few representatives. The main idea is to summarize the training data and to identify few prototypes that can capture the variations of the signatures within each class. These variations could be due to different mine types, different soil conditions, different weather conditions, etc. Fuzzy memberships are assigned to the representatives to capture their degrees of sharing among the mine and clutter classes.

The rest of this paper is organized as follows. Section 2 gives an overview of the GPR data and the LMS detector. Section 3 describes the different steps of the proposed detection system. The experimental results are presented in section 4, and concluding remarks are given in section 5.

II. ANOMALY DETECTION

In this section, we present a brief description of the GPR data, the pre-processing steps, and the LMS pre-screener. A more detailed description of these steps can be found in [15], [14].

A. GPR Data

The input data consists of a sequence of raw GPR signatures sampled by vehicle-mounted antennas as it travels

Hichem Frigui is with the Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY 40292, USA (email: h.frigui@louisville.edu).

Paul Gader is with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA (email: pgader@cise.ufl.edu).

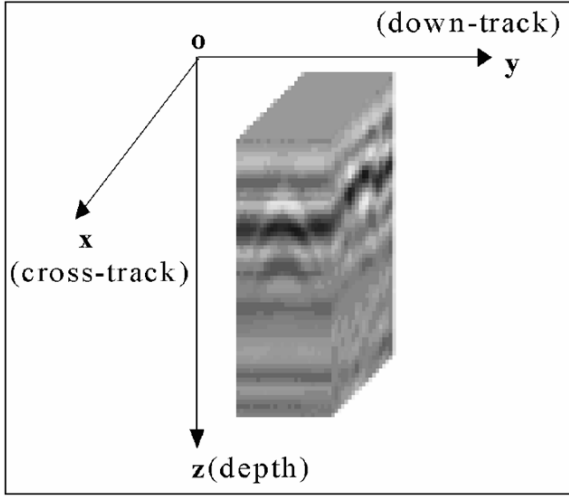


Fig. 1. a collection of few GPR scans

forward. The Wichmann GPR of NIITEK is used to collect 24 channels of data. Adjacent channels are spaced approximately 5 centimeters apart in the cross-track direction, and sequences (or scans) are taken at approximately 5 centimeter down-track intervals. The sequence at each cross-track and down-track position contains 416 time samples (which are approximately related to depth) at which the GPR signal return is reported. The collected input data is represented by a 3-Dimensional matrix of sample values, $S(z, x, y)$, $z = 1, \dots, 416$; $x = 1, \dots, 24$; $y = 1, \dots, N_S$, where N_S is the total number of collected scans, and the indices z, x , and y represent depth, cross-track position, and down-track positions respectively. A collection of scans, forming a volume of data, is illustrated in Fig. 1.

B. Pre-processing and the LMS Pre-screener

First, we identify the location of the ground bounce as the signal's peak and align the multiple signals with respect to their peaks. This alignment is necessary because the vehicle-mounted system cannot maintain the radar antenna at a fixed distance above the ground. The top part of each signal, up to few samples beyond the ground bounce are discarded. The remaining signal samples are divided into N depth bins, and each bin would be processed independently. The reason for this segmentation is to compensate for the high contrast between the responses from deeply buried and shallow anomalies. Next, an adaptive LMS is applied to the energy at each depth bin. The LMS assigns a confidence value to each point in the cross-track, down-track plane based on its contrast with a neighboring region. The components that satisfy empirically pre-determined conditions are considered as potential targets. Their cross-track x_s , and down-track y_s positions of the connected component center are reported as alarm positions for further processing.

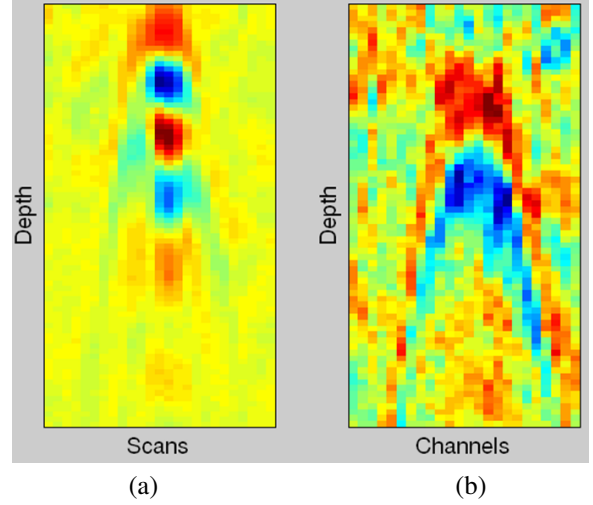


Fig. 2. (a) (depth-downtrack), and (b) (depth,cross-track) views of a sample mine signature

III. FEATURE-BASED LAND MINE DETECTION

A. Edge Histogram Descriptor

We use a variation of the MPEG-7 Edge Histogram Descriptor (EHD) [16] as a feature representation of the GPR signatures. The basic EHD has undergone rigorous testing and development, and thus, represents one of the mature and generic texture descriptors. For a generic image, the EHD represents the frequency and the directionality of the brightness changes in the image. Simple edge detector operators are used to identify edges and group them into five categories: vertical, horizontal, 45° diagonal, 135° diagonal, and isotropic (non-edges). The EHD would include five bins corresponding to the above categories.

For the GPR data, we adapt the EHD to capture the spatial distribution of the edges within a 3-D GPR data volume. To keep the computation simple, we still use 2-D edge operators, but we compute two types of edge histograms. The first one is obtained by fixing the cross-track dimension and extracting edges in the (depth, down-track) plane. The second edge histogram is obtained by fixing the down-track dimension and extracting edges in the (depth, cross-track) plane. Fig. 2 displays a (depth,down-track) plane and a (depth,cross-track) plane of a sample mine signature. As it can be seen, the edges in these planes and their spatial distribution constitute an important feature to characterize the mine signatures.

Let $S_{zy}^{(x)}$ be the x^{th} plane of the 3-D signature $S(x, y, z)$. First, for each $S_{zy}^{(x)}$, we compute four categories of edge strengths: vertical, horizontal, 45° diagonal, and 135° diagonal. If the maximum of the edge strengths exceeds a certain preset threshold, θ_G , the corresponding pixels is considered to be an edge pixel. Otherwise, it is considered a non edge pixel. Next, each $S_{zy}^{(x)}$ image is vertically subdivided into 4 overlapping sub-images $S_{zy_i}^{(x)}$, $i = 1, \dots, 4$. For each $S_{zy_i}^{(x)}$, we compute a 5 bin edge histogram, $H_{zy_i}^{(x)}$, where the bins correspond to the 4 edge categories, and the non-edge pixels. The down-track component of the EHD, EHD^d is defined as

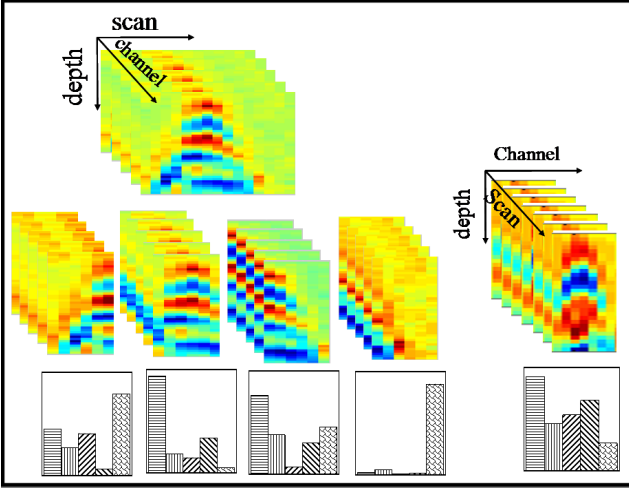


Fig. 3. Extraction of the EHD for a 3-D mine signature

the concatenation of 4 five-bin histograms:

$$EHD^d(S_{xyz}) = [\bar{H}_{zy_1} \bar{H}_{zy_2} \bar{H}_{zy_3} \bar{H}_{zy_4}], \quad (1)$$

where \bar{H}_{zy_i} is the cross-track average of the edge histograms of sub-image $S_{zy_i}^{(x)}$ over N_C channels, i.e.,

$$\bar{H}_{zy_i} = \frac{1}{N_C} \sum_{x=1}^{N_C} H_{zy_i}^{(x)}.$$

To compute the cross-track component of the EHD, EHD^x , we fix the scans, and compute the 4 edge strengths on the $S_{zx}^y, y = 1, \dots, N_S$ (depth, cross-track) planes. Since these planes do not have enough columns (typically < 7), they are not divided into sub-images, and only one global histogram per plane, H_{zx}^y , is computed. Then, EHD^x is computed as the down-track average of the edge histograms over N_S scans, i.e.,

$$EHD^x(S_{xyz}) = \frac{1}{N_S} \sum_{y=1}^{N_S} H_{zx}^{(y)} \quad (2)$$

The EHD of each 3-D GPR alarm is a 25-D histogram that concatenates the down-track and cross-track EHD components, i.e.,

$$EHD(S_{xyz}) = [EHD^y(S_{xyz}) \ EHD^x(S_{xyz})]. \quad (3)$$

The extraction of the EHD is illustrated in Fig. 3

B. Training Signatures

The training data consists of a set of alarms reported by the LMS pre-screener and labeled as mines or false alarms using the ground truth. The LMS reports the cross-track (x_s) and down-track (y_s) position (center of connected component) of each alarm s . Since the ground truth for the depth value (z_s) is not provided, we visually inspect all mine signatures and estimate this value. For the false alarms, this process is not trivial as false alarms can have different characteristics and their signature can extend over a different number of

samples. Instead, for each reported false alarm, we extract five equally spaced depths (z_{s_1}, \dots, z_{s_5}) covering the entire depth range.

Each signature s consists of a 30 (depth values) by 15 (scans) by 7 (channels) volume extracted from 7 consecutive channels extracted from channel x_s of the aligned GPR data and centered at (y_s, z_s) .

C. Clustering the Training Signatures

The signatures within each class are expected to exhibit significant variations. For instance, clutter signatures can be caused by different types of buried objects. Similarly, mine signatures can have multiple subclasses corresponding to mines of different types and sizes, mines buried at different depths, different soil and weather conditions, etc. To reduce the size of the training samples and identify few representatives that can capture these within-class variations, we use the self-organizing feature maps (SOFM) [17] to cluster the mine and false alarms signatures separately. We will refer to the clusters' representatives (R_i) as prototypes. We use R_i^M to denote the prototypes of the mine signatures, and R_i^C to denote the prototypes of the clutter signatures.

For further processing, each prototype, R_i , is assigned a fuzzy membership in the class of mines, $u^M(R_i)$, and a fuzzy membership in the class of false alarms $u^C(R_i)$. We use a minimum distance and a Fuzzy C-Means [18] based labeling. Specifically, for each R_i , we identify the closest mine prototype R_i^M and the closest clutter prototype R_i^C , and assign a label using

$$u^M(R_i) = \frac{1/\text{dist}(R_i, R_i^M)}{1/\text{dist}(R_i, R_i^M) + 1/\text{dist}(R_i, R_i^C)} \quad (4)$$

D. Fuzzy K-NN based confidence assignment

Each potential target (identified by the pre-screener) is tested at multiple depth values. We slide a $30 \times 15 \times 7$ window size along the depth axis with a 50% overlap between 2 consecutive signatures. A maximum of 10 signatures are extracted for each target. For each signature, we compute the EHD, and use A fuzzy K-NN [19] based rule to assign a confidence value. Then, the 10 confidence values are combined using an order statistics (OS) operator [20] to generate a single confidence value.

TO compute the confidence value for a given test signature, S_T , we compute its distance to all representative prototypes. Then we sort these distances, and identify the top K nearest neighbors S_T^1, \dots, S_T^K . We experimented with two fuzzy versions of the K-NN. In the first one, we compute the confidence values using

$$\text{Conf}(S_T) = \frac{\sum_{k=1}^K u^M(S_T^k) \times \frac{1}{\text{dist}(S_T, S_T^k)}}{\sum_{k=1}^K 1/\text{dist}(S_T, S_T^k)}. \quad (5)$$

In this version, the confidence value depends on the relative distances of the K nearest neighbors. Relatively close prototypes will contribute more to the overall confidence value. In

TABLE I
NUMBER OF METAL AND PLASTIC CASED MINES AND MINE
SIMULANTS AND THEIR BURIAL DEPTHS.

	Depth								Total
	-1"	0"	1"	2"	3"	4"	5"	6"	
Metal	12	48	42	121	43	101	4	53	424
Plastic	6	21	8	57	29	24	0	58	203
Simulants	0	0	0	37	18	26	0	0	81

the second version of the K-NN, we compute the confidence value using

$$Conf(S_T) = \sum_{k=1}^K u^M(S_T^k) \times \frac{1}{1 + \max\left(0, \frac{dist(S_T, S_T^k) - \bar{D}}{\eta}\right)} \quad (6)$$

In (6), the \bar{D} and η parameters are determined experimentally using the training data. Eq. 6 can be considered a possibilistic version of the K-NN, where the overall confidence value depends on the absolute distance of the nearest neighbors to the prototypes. Test signatures that are far from all prototypes, will be assigned low confidence values. This is not the case if eq. (5) is used.

IV. EXPERIMENTAL RESULTS

The EHD based detector was developed and tested on GPR data collected from outdoor test lanes at three different locations. The first two locations, site 1 and site 2, were temperate regions with significant rainfall, whereas the third collection, site 3, was a desert region. The lanes are simulated roads with known mine locations. Lanes at site 1 are labeled lanes 1, 3, and 4, and are 500 meters long and 3 meters wide. Lanes at site 2 are labeled lanes 3, 4, 13, 14, and 19, and are 50 to 250 meters long and 3 meters wide. Lanes at site 3 are labeled lanes 51 and 52, and are 300 meters long and 3 meters wide. All mines are Anti-Tank (AT) mines. Multiple data collections were performed at each site at different dates resulting in a total of 708 mine encounters. The number, type, and burial depth of the mines are given in table I. For all the 10 lanes in the 3 collections, the LMS has identified a total of 1777 alarms.

The identified alarms were used to train and test the EHD detector. We use a lane-based cross validations: We use the alarms of 9 lanes to train and test on one lane. This process would be repeated 10 times so that each lane is tested once. For each cross validation, training alarms from 9 lanes would be partitioned into mine and clutter using the available ground truth. Then, the self-organizing feature maps (SOFM) [?] would be used to cluster each group of signatures into a 10×10 map to identify the representative prototypes. Fig. 4 displays the SOM map of the mine prototypes for one of the 10 cross validation sets. This map, which includes 97 mine prototypes, represents a summary of about 600 mine signatures. As it can be seen, some of the prototypes have strong and well-structured signatures, while others have weak signatures. The fuzzy labels that are assigned to these prototypes (see eq. (4)) would quantify this variation.

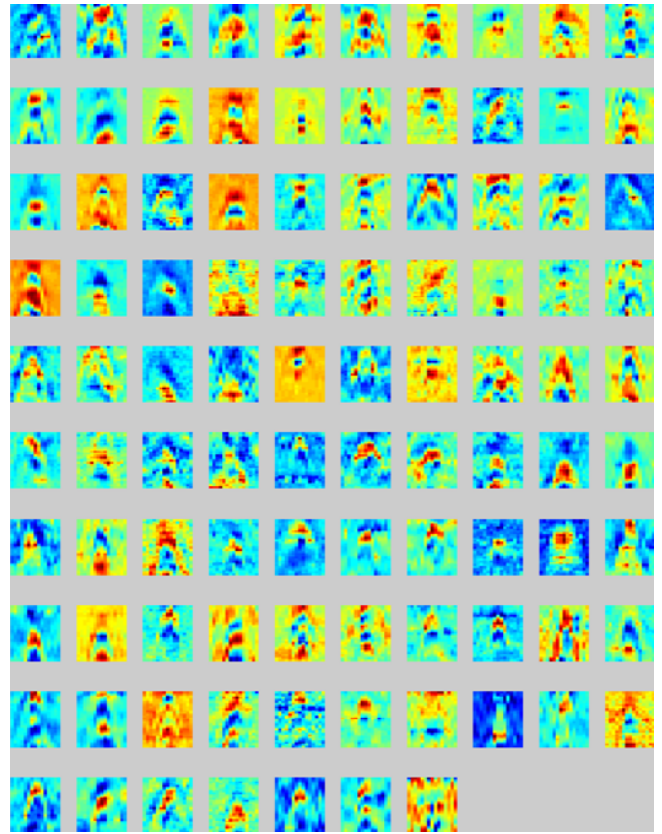


Fig. 4. SOM map of the mine prototype signatures for one training set.

Fig. 5 displays the SOM map of the clutter prototypes for the same cross validation sets. This map includes 100 prototypes and represents the summary of about 1000 clutter signatures. As it can be seen, some of the clutter prototypes (e.g. bottom left corner) resemble the signatures of weak mine. These prototypes will be assigned low membership values in the class of mines and would contribute to the overall confidence value (see eq. (5)). In other words, clutter signatures that have partial edge structure would be treated differently from clutter signatures that have high energy but no structure.

We have experimented with the two K-NN versions, and we have found that, in general, equations (5) and (6) yield comparable performance. However, there are few cases where the test signature (usually clutter) is not similar to any of the identified prototypes. In this case, the possibilistic K-NN outperforms the fuzzy K-NN.

The EHD detector was scored in terms of Probability of Detection (PD) vs. False Alarm Rate (FAR). Confidence values were thresholded at different levels to produce Receiver Operating Characteristic (ROC) curve. For a given threshold, a mine is detected if there is an alarm within 0.25 meters from the edge of the mine with confidence value above the threshold. Given a threshold, the PD is defined to be the number of mines detected divided by the number of mines. The FAR is defined as the number of false alarms per square

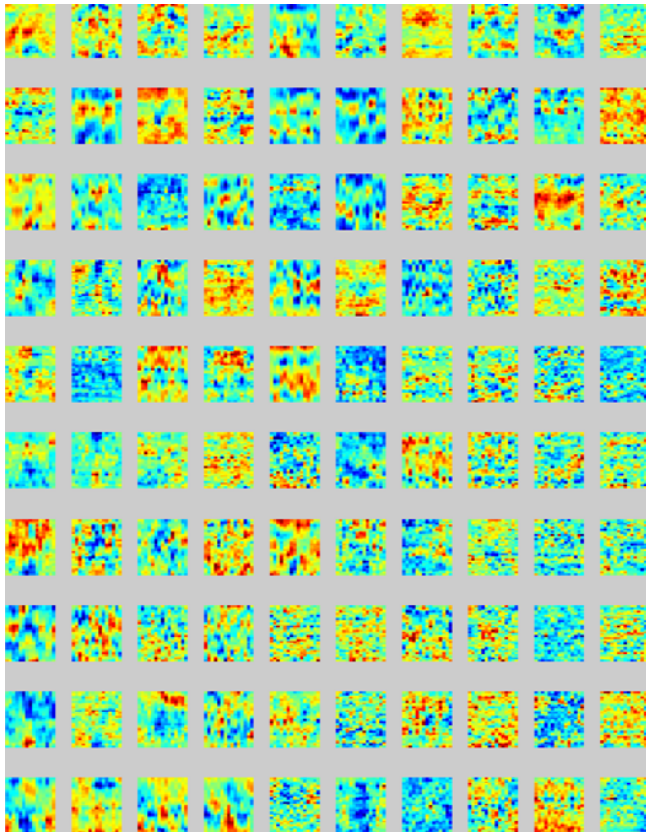


Fig. 5. SOM map of the clutter prototype signatures for one training set.

meter.

The results of the EHD detector are compared with those of the pre-screener and with the results obtained using the HMM detector [11], [12]. Fig. 6 shows the ROCs for all the passes of the 3 collections. The ROCs are displayed for the confidence values generated by the LMS prescreener and the EHD detector. As it can be seen, when compared to the LMS ROC, the EHD ROC is shifted left (i.e., lower FAR for the same PD), and shifted up (higher PD for the same FAR). Thus, one can conclude that the EHD detector can discriminate between the mine and clutter signatures identified by the prescreener. In fact, examination of the confidence values of individual alarms has indicated that the EHD has increased the confidence values of several "weak" mine signatures considerably. Fig. 7 displays samples of these mine signatures. Similarly, the EHD has reduced the confidence values of several clutter signatures significantly. These are usually signatures with high energy content that don't have the coherent spatial edge distribution. Fig. 8 displays samples of these clutter signatures.

V. CONCLUSION

In this paper, we have proposed an approach for land mine detection based on edge histogram descriptor and fuzzy K-nearest neighbors. In addition to being simple and efficient, our approach is data driven and thus, could be easily re-

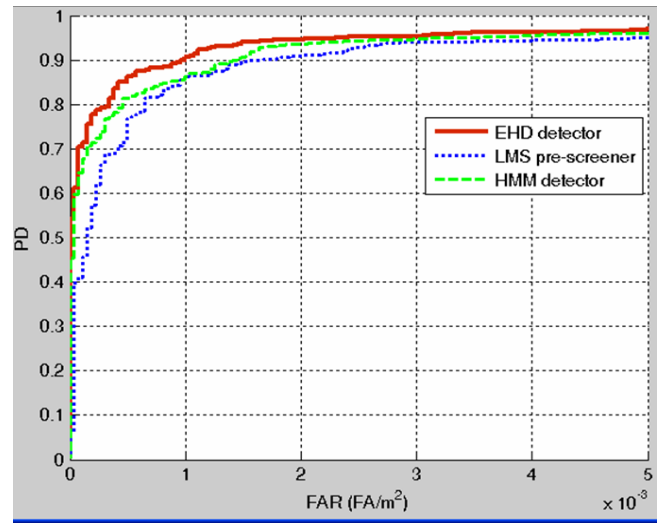


Fig. 6. Comparison of the EHD, LMS, and HMM ROC's

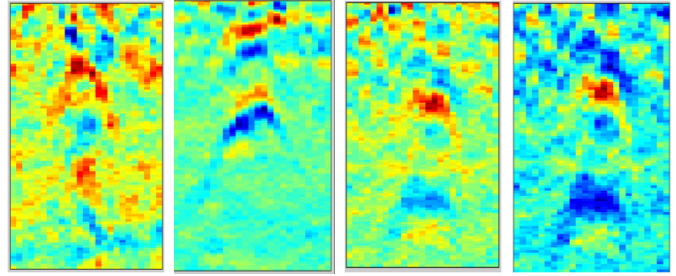


Fig. 7. Sample mine signatures where the EHD has increased the confidence values significantly.

trained and adapted to data collected from other sites and/or with different GPR sensors. The fuzzy labels assigned to the mine and false alarm representatives help the system assign soft confidence values that can reflect the ambiguity of the signatures. This feature is important if the results of the EHD detector are to be fused with those obtained by different classifiers. The ROC on data collected from several lanes at different sites show that the EHD algorithm can reject several false alarms identified by LMS without affecting the detection rate.

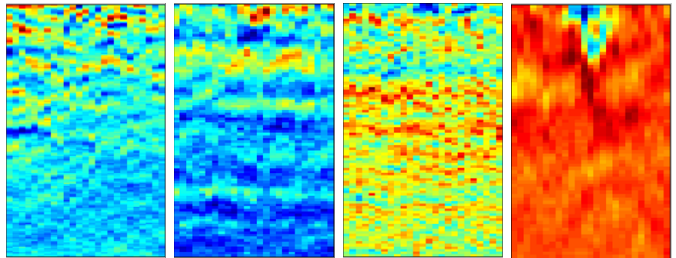


Fig. 8. Sample clutter signatures where the EHD has reduced the confidence values significantly.

ACKNOWLEDGMENT

This work was supported in part by a U.S. Army Grant Number DAAB15-02-D-0003, and an Office of Naval Research award number N00014-05-10788 for H. Frigui, and by the U. S. Army Research Office and U. S. Army Research Laboratory under Cooperative Agreement Number DAAD19-02-2-0012 for P. Gader. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Office of Naval Research, Army Research Laboratory, or the U. S. Government.

REFERENCES

- [1] "Landmines, mine action news from the united nations, vol. 3.2, fourth quarter 1998," .
- [2] "Hidden killers: The global landmine crisis, united states department of state report, publication no. 10575, sept. 1998," .
- [3] T. R. Witten, "Present state of the art in ground-penetrating radars for mine detection," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando FL, 1998, pp. 576–586.
- [4] P. D. Gader, H. Frigui, B. Nelson, G. Vaillette, and J. M. Keller, "New results in fuzzy set based detection of landmines with gpr," in *Detection and Remediation Technologies for Mines and Minelike Targets IV*, Orlando FL, 1999, pp. 1075–1084.
- [5] D. Carevic, "Clutter reduction and target detection in ground penetrating radar data using wavelets," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets IV*, Orlando FL, 1999, pp. 973–978.
- [6] D. Carevic, "Kalman filter-based approach to target detection and target-background separation in ground-penetrating radar data," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets IV*, Orlando FL, 1999, pp. 1284–1288.
- [7] A. Gunatilaka and B. A. Baertlein, "Subspace decomposition technique to improve GPR imaging of anti-personnel mines," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando FL, 200.
- [8] H. Brunzell, "Detection of shallowly burried objects using impulse radar," *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, pp. 875–886, 1999.
- [9] S. Yu, R. K. Mehra, and T. R. Witten, "Automatic mine detection based on ground penetrating radar," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets IV*, Orlando FL, 1999, pp. 961–972.
- [10] H. Frigui, K. satyanarayana, and P. Gader, "Detection of land mines using fuzzy and possibilistic membership functions," in *proceedings of the IEEE Conference on Fuzzy Systems*, Saint Louis, Missouri, 2003.
- [11] P. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden markov models," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 1231–1244, 2001.
- [12] H. Frigui, K. C. Ho, and P. Gader, "Real-time land mine detection with ground penetrating radar using discriminative and adaptive hidden markov models," *EURASIP Journal on Applied Signal Processing*, vol. 12, pp. 1867–1885, 2005.
- [13] P. Gader, B. Nelson, H. Frigui, G. Vaillette, and J. Keller, "Fuzzy logic detection of landmines with ground penetrating radar," *Signal Processing, special issue on fuzzy logic in signal processing*, vol. 80, pp. 1069–1084, 2000.
- [14] P. Gader, W. H. Lee, and J. N. Wilson, "Detecting landmines with ground penetrating radar using feature-based rules, order statistics, and adaptive whitening," (accepted for publication).
- [15] P. A. Torrione, C. S. Throckmorton, and L. M. Collins, "Performance of an adaptive feature-based processor for a wideband ground penetrating radar system," (in press).
- [16] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG 7: Multimedia Content Description Language*, John Wiley, 2002.
- [17] T. Kohonen, *Self-Organization and Associative Memory*, Springer Verlag, 1989.
- [18] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [19] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.
- [20] P. Gader, R. Grandhi, W. Lee, J. Wislon, and D. Ho, "Feature analysis for the NIITEK ground-penetrating radar using order-weighted averaging operators for landmine detection," in *SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets*, 2004.

An Investigation of Using the Spectral Characteristics From Ground Penetrating Radar for Landmine/Clutter Discrimination

K. C. Ho, *Senior Member, IEEE*, Lawrence Carin, *Fellow, IEEE*,
Paul D. Gader, *Senior Member, IEEE*, and Joseph N. Wilson, *Member, IEEE*

Abstract—Ground penetrating radar (GPR)-based discrimination of landmines from clutter is known to be challenging due to the wide variability of possible clutter (e.g., rocks, roots, and general soil heterogeneity). This paper discusses the use of GPR frequency-domain spectral features to improve the detection of weak-scattering plastic mines and to reduce the number of false alarms resulting from clutter. The motivation for this approach comes from the fact that landmine targets and clutter objects often have different shapes and/or composition, yielding different energy density spectrum (EDS) that may be exploited for their discrimination (this information is also present in time-domain data, but in the frequency domain we can remove a phase if desired and can reveal better spatial characteristics and therefore often achieve greater robustness). This paper first applies the finite-difference time-domain (FDTD) modeling technique to establish the theoretical foundation. The method to generate EDS from GPR measurements is then described. The consistency of the frequency-domain features is examined through two different GPRs that have different spatial sampling rates and frequency bandwidths. Experimental results from several test sites, based on GPR data collected over buried mines and emplaced buried clutter objects, corroborate the theoretical development and the effectiveness of the proposed spectral feature to increase the accuracy of landmine detection and discrimination.

Index Terms—Energy density spectrum (EDS), finite-difference time-domain (FDTD) modeling, ground penetrating radar (GPR), landmine detection.

I. INTRODUCTION

LANDMINE detection has been the subject of several investigations over the past few years [1]–[36]. The research is driven not only by need in military operations but also for humanitarian purposes to clean up minefields left after wars (minefields are responsible for more than 30 000 deaths and injuries every year).

Manuscript received May 22, 2006; revised June 24, 2007. This work was supported in part by the U.S. Army RDECOM CERDEC NVESD under Contracts DAAB15-01-D-0004/DO#0002, DAAB15-02-D0003/DO#0001, and DAAB15-02-D0003/DO#0003.

K. C. Ho is with the Department of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211 USA (e-mail: hod@missouri.edu).

L. Carin is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA.

P. D. Gader and J. N. Wilson are with the Department of Computer Information and Science Engineering, University of Florida, FL 32611 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2008.915747

Because landmines are often buried underground, landmine detection relies on ground-penetrating sensors to capture the signal response. Perhaps the most popular sensor for landmine detection is electromagnetic induction, often termed a metal detector (MD) [8]–[14]. If the landmine casing contains significant metal, it will typically trigger responses in the MD and be detected (the MD response drops off as $1/r^6$ where r is the target-sensor distance, thus MDs often have difficulty with low-metal-content mines at significant depths). The MD is also significantly impacted by the quantity of metal in the target, with this a significant problem for low-metal-content plastic mines. Metal detectors also provide limited discrimination capability, and therefore they suffer from false alarms due to ubiquitous metal clutter. Many currently developed landmines are either made of plastic or have very low metal content. As a result, an MD alone is not able to achieve a high probability of detection with a correspondingly low probability of false alarm, and additional sensors are needed. We note that radars are sensitive to plastic mines if there is sufficient contrast between the dielectric properties of the mine and the soil. Moreover, radar signatures fall off as $1/r^4$.

Ground penetrating radar (GPR) is a sensor modality that has recently witnessed improved classification performance for landmine detection [15]–[36]. This improvement in performance has been manifested by improved electronics (e.g., wider bandwidth and better antennas) and enhanced signal-processing architectures. GPRs may operate in the time or frequency domains. One must balance the desire for significant ground penetration ability, which necessitates low frequencies, with the desire for spatial resolution, which requires wider bandwidths. Many current systems operate from a lower frequency of approximately 0.5 GHz to upper frequencies approaching 10 GHz.

The GPR signal from a landmine is dependent on the mine's size, shape, and composition, as well as its burial depth and orientation. In addition to the properties of the mine itself, electrical characteristics of the soil also play an important role on the signature of landmines and clutter. For example, if the dielectric constant of the mine and soil are similar, the electrical discontinuity manifested by the mine-soil heterogeneity may be small, yielding a weak landmine signature. To address this problem, one may lower the detection threshold, thereby increasing the probability of detecting mines with weak signatures; however, this typically will cause a significant increase in the number of false alarms.

Rather than simply using the (often weak) signature amplitude to perform detection of landmines, one may consider exploiting the spectral properties of the signature to use potential mine-specific features. In our paper, we perform classification based on spectral features extracted from the entire GPR waveform signature. The rationale to exploit the spectral characteristics for classification is that landmine targets and clutter objects often have different shapes as well as composition, which yields different amounts of energy return at different frequencies, and hence different energy density spectra. It is well known to the electromagnetics community that the entire scattered waveform (A-scan) from a target illuminated by an ultrawideband pulse conveys signature information. Particularly for stationary landmine targets in clutter environment, it is essential that no signature information is excluded. We therefore apply the entire signature waveform when generating the spectral characteristics of a target.

As indicated above, the A-scan signature waveform of a landmine (or general target) is characteristic of the target itself and is source independent (although the strength of the radar return may vary with a changing source). However, one must view the landmine and surrounding soil medium as a composite target. For a fixed landmine, the characteristics of the signature waveform change with variable surrounding soil properties and for variable mine positions (e.g., depth and orientation). Depending on orientation, the target looks very different, and one target may look like another if the orientations are changed. This may present a significant challenge, due to changing soil properties with time and space and due to different target burial properties. To examine the significance of this issue in detail, we perform numerical simulations with a three-dimensional finite-difference time-domain (FDTD) numerical model [37]. The accuracy of the FDTD model is first validated by comparing it to measured data from actual plastic mines. FDTD is subsequently employed to examine the spectral characteristics of mines as a function of target depth and soil properties. Based on the insight accrued from this modeling, we observe that the spectral signature is relatively robust to changing environmental conditions, motivating its use subsequently for landmine detection. We use here the energy density spectrum (EDS) to obtain the spectral characteristics. Its features are then deduced from EDS to improve landmine detection and clutter discrimination. Furthermore, fusion results with time-domain features are also provided to demonstrate the advantages and usefulness of the proposed EDS spectral feature technique. Extensive experimental results corroborate that the proposed EDS technique significantly improves the detection performance of landmines, especially in the presence of various clutter objects such as pieces of woods, rocks, plastics, and metal debris.

We have searched through the literature and have not found any previous work on using the spectral characteristics from GPR measurements over a target for landmine and clutter discrimination. Some researchers have investigated the use of complex natural resonance in the GPR late-time response for the classification of unexploded ordnances (UXOs) [38]–[40]. The complex resonance frequencies are estimated from the late-time response using the parametric estimation technique, and the estimated resonance frequencies are used for UXO

classification. Our work focuses on landmine and clutter discrimination, using a different approach and methodology. In particular, we use the entire time response from the GPR measurement instead of the late-time response only to create the EDS. Furthermore, the proposed technique uses the shape in the EDS between landmine and clutter objects for discrimination, and it does not estimate resonance frequencies.

The paper is organized as follows. Section II discusses the FDTD technique to model plastic mine targets and derives their theoretical spectral characteristics. Section III presents procedures to generate EDS from GPR data measurements. Section IV contains the experimental results using the data with buried landmines as well as clutter objects that are collected from several test sites at different geographic locations with different soil types. The conclusions from this study are summarized in Section V.

II. FDTD MODELING

This section applies the FDTD modeling of weak-scattering plastic landmines. The FDTD helps us understand the phenomenology that produces the distinct spectral characteristics of some weak-scattering plastic landmines and provides a good tool for us to analyze how the spectral characteristics are affected by the background-soil electrical parameters such as soil conductivity and dielectric constant. It also assists in the design of the spectral mask for the proposed algorithm for landmine and clutter discrimination.

The FDTD is a widely applied electromagnetic modeling tool appropriate for analyzing general three-dimensional scattering and radiation problems [37]. We apply it here to synthesize the electromagnetic signature of three-dimensional buried landmines. The antenna system used in the simulations is similar to that investigated in [37] and [41], and therefore no further details of the antennas are provided here. Measurements were performed with a time-domain GPR system operating over the 0.5–8 GHz frequency band, with a design analogous to that in [37] and [41]. For a system with such a wide bandwidth, it is essential to model the detailed internal components of a plastic landmine (this is obviously not important for metal-cased mines, for which there is little, if any, electromagnetic penetration). To perform such modeling, we referred to *Jane's Ammunition Handbook* for the characteristics of mines [42]. *Jane's* gives cross-sectional dimensions of the internal components of landmines, as well as photographs. Using these data, and knowledge of the electromagnetic properties of typical plastics, we approximated the internal components of the landmine within the FDTD model. In the subsequent discussion, we do not give the name of the explicit mines considered in these studies for security reasons. However, we do characterize the general mine properties (e.g., dimensions). Details on many such mines may be found in [42].

The first curve in Fig. 1 is the FDTD-computed frequency-domain signature (magnitude) for a moderately size circular plastic antitank mine (height: 11.5 cm, diameter: 23 cm). Frequency-domain signatures are also shown in Fig. 1 and were obtained from measured data that were collected at a test facility at a temperate site. The mine in Fig. 1 was buried at a depth of approximately 10 cm from the top of the mine

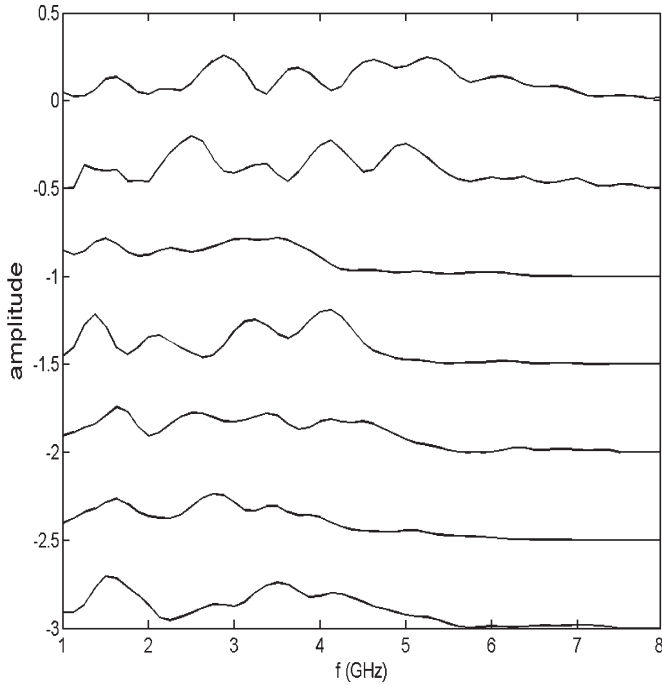


Fig. 1. Spectral characteristics of a plastic landmine. The top curve is from FDTD modeling, and the rest of the curves are from measurements.

to the soil interface, and the soil properties are $\varepsilon_r = 3$ and $\sigma = 0.05$ S/m from on-site measurement. The six measured frequency-domain signatures come from the same type of mine but are different and therefore have variation.

The results for another mine type is presented in Fig. 2, where the first curve is from FDTD modeling, and the rest of the curves are from data measurements. This is a relatively large plastic antitank mine (height: 7 cm, diameter: 31 cm), and the top of the mine was buried flush to the soil interface. The soil properties in this case are $\varepsilon_r = 5.5$ and $\sigma = 0.01$ S/m via on-site measurement when the data were collected. Again, the six curves from the data measurement came from the same type of mine but from different mines, which contributes to their variation. In the results presented in Figs. 1 and 2, the sensor is situated above the center axis of the mine (the end of the antenna is 5 cm from the interface). The curves in the two figures are translated with -0.5 decrements for purposes of comparison, and the absolute scale in the y -axis does not have meaning.

The comparison in Figs. 1 and 2 is typical of what we have observed from field data for actual plastic landmines (these are not “sandbox” laboratory measurements but rather GPR measurements of actual mines emplaced in test lanes over many years). In the two figures, the theoretical spectrum is on the top and the rests are measurements. The theoretical results match the measurement results very well. Note that both the measured and computed data are characterized by peaks in the frequency-domain magnitude spectra. These spectral peaks are attributed to the reflective scattering from the target. When Fig. 1 is examined carefully, it is realized that the measured frequency-domain signatures vary significantly in the high-frequency region above 3 GHz. However, the signatures have more consistency in the region between 1 to slightly above 2 GHz. We have a similar observation for Fig. 2.

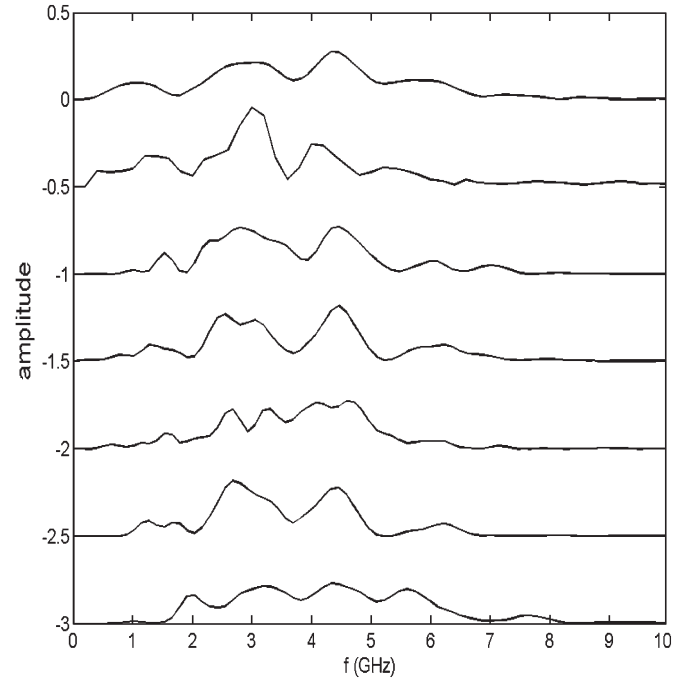


Fig. 2. Spectral characteristics of another plastic landmine. The top curve is from FDTD modeling, and the rest of the curves are from measurements.

To compare the theoretical signature with the measured ones, we use the following metric correlation coefficient:

$$\rho_{AB} = \frac{\sum_i A(f_i)B(f_i)}{\sqrt{\sum_i A(f_i)^2} \sqrt{\sum_i B(f_i)^2}}$$

where $A(f_i)$ and $B(f_i)$ are the two frequency-domain signatures to be compared. The correlation coefficient is equivalent to the mean-square error measure when the two frequency domain signatures have been normalized to unity energy. The correlation coefficient allows us to compare the shape of the two signatures and ignores the effect in the returned energy strength. The correlation coefficient is between -1 and 1 . The closer its value to one, the higher the similarity between the two signatures $A(f_i)$ and $B(f_i)$.

The correlation coefficient values when setting $A(f_i)$ to be the theoretical one and $B(f_i)$ to be the measured signatures are $[0.94, 0.93, 0.90, 0.92, 0.91, 0.93]$ for the signatures in Fig. 1, and are $[0.90, 0.94, 0.95, 0.93, 0.93, 0.93]$ for the signatures in Fig. 2. These values are very close to unity, indicating high similarity between the theoretical model and the measurements.

The two mine types in Figs. 1 and 2 show some different spectral characteristics, which motivates us to consider the use of spectral features to design a classifier. One feature to be noted is that when using EDS for landmine classification, the GPR should have a high-enough frequency resolution to provide a high-quality spectrum of a target. Furthermore, the measured EDS from one mine type should have a much smaller variance compared to the difference in the EDS between two different mine types. This dictates a very-well-controlled GPR measurement and a very stable environmental condition that are often not achievable in practice. However, as illustrated toward the end of Section III, our study finds that the EDS between landmine and some clutter objects have very large differences

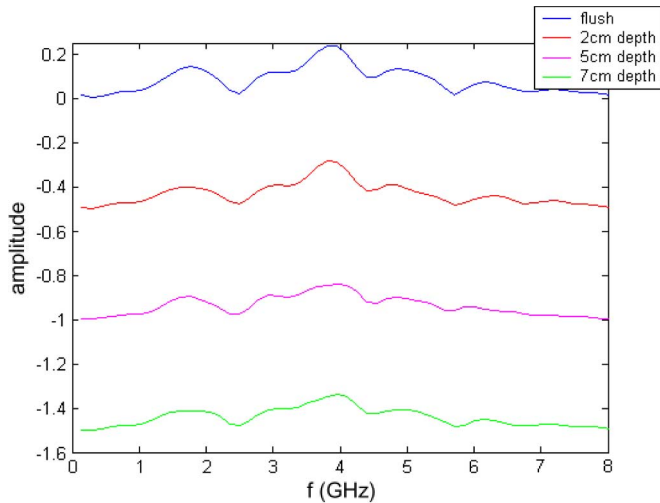


Fig. 3. Spectral characteristics of the plastic landmine as shown in Fig. 1 buried at different depths.

(small correlation coefficients). As a result, the EDS is found to be quite useful for providing some discrimination ability between landmine and clutter objects.

Having demonstrated the accuracy of the three-dimensional FDTD model by comparing it to measured data, we now exercise the model to perform studies that would be difficult to replicate experimentally. All results presented in the subsequent discussion are for the landmine considered in Fig. 1, and the results are representative of results we have found for numerous buried plastic landmines. In Fig. 3, we consider the same mine and soil properties as considered in Fig. 1, but now results are presented as a function of target depth (as measured from the top of the mine). The sensor is situated above the center axis of the mine when generating the frequency spectra. The spectra in Fig. 3 are shifted vertically so that they can be compared with each other, and the absolute scale in the y -axis does not have meaning. We observe in Fig. 3 that the frequency-dependent signature of the mine (amplitude) is relatively insensitive to the target depth, for fixed soil properties. In fact, the correlation coefficients for the spectra in Fig. 3 are $\rho_{12} = 0.99$, $\rho_{13} = 0.96$, $\rho_{14} = 0.95$, $\rho_{23} = 0.97$, $\rho_{24} = 0.96$, and $\rho_{34} = 0.99$, where the spectra are numbered in the order as they appear. Although the composite mine-soil target changes with variable depth, the properties of the soil surrounding the mine do not (only the distance from the mine to the interface changes). The spectral peaks associated with a plastic mine may be attributed principally to multiple reverberant scattering within the mine itself and are apparently not sensitive to the surrounding soil properties, although the surrounding soil will change the reflection intensity. We now examine the effect of background-soil electrical parameters on the spectral characteristics of plastic landmine. In Fig. 4, we present results for the same plastic antitank landmine as considered in Fig. 3, buried with the top of the mine flush with the soil interface and the sensor centered over the mine as discussed above. Again, the spectra in Fig. 4 are shifted vertically for the purpose of comparison, and the absolute scale in the y -axis does not have meaning. In these examples the soil conductivity is fixed at $\sigma = 0.05$ S/m and the dielectric constant is varied to three values: $\epsilon_r = 2.5$,

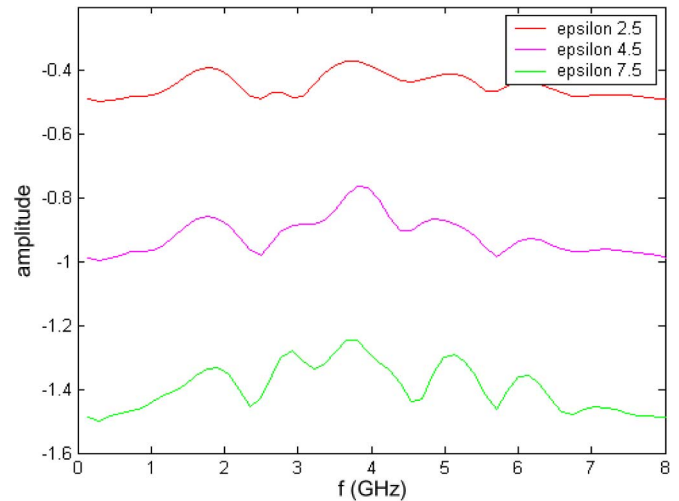


Fig. 4. Spectral characteristics of the plastic landmine as shown in Fig. 1 with different dielectric constants in the soil.

$\epsilon_r = 4.5$, and $\epsilon_r = 7.5$. The internal components of the mine are composed principally of plastic ($\epsilon_r = 2.5$) and air pockets, and typically the dielectric constant of the explosive is close to that of a plastic. Therefore, there is substantial variation in the electrical contrast between the mine and soil for soil permittivity $\epsilon_r = 7.5$ and far less contrast for $\epsilon_r = 2.5$. The correlation coefficients of the curves are found to be $\rho_{12} = 0.97$, $\rho_{13} = 0.94$, and $\rho_{23} = 0.97$, where the curves are numbered in the order as they appear. The correlation coefficients values are quite close to one, indicating a high similarity among them.

We observe from Fig. 4 that the spectral properties of the landmine vary as a function of changing soil properties. The variation in the spectra that is above 2.5 GHz is quite significant as the dielectric constant ϵ increases. The variation in the spectra is less, or the spectra are more stable when the frequency is below 2.5 GHz. This observation also appears in Fig. 3 as the depth of the landmine increases. Consequently, when using the spectra to improve the detection of a weak-scattering landmine, more emphasis should be placed on the frequency region below 2.5 GHz. Also, the shape of the spectra would be preferable to the spectral peak frequencies as the spectral peak frequencies tend to vary significantly.

Finally, we note that in Fig. 4 when comparing the spectrum with soil $\epsilon_r = 2.5$, the spectrum with soil properties $\epsilon_r = 7.5$ is more different than that for $\epsilon_r = 4.5$. Consequently, one would expect that the spectrum should deviate even further from the $\epsilon_r = 2.5$ case as the dielectric constant of the soil increases, for example, to $\epsilon_r = 15$. In this case, one may not expect a single spectrum to cover all soil conditions. In our work, we have not seen many cases for which the soil was characterized by $\epsilon_r = 15$. In such cases, we would most likely condition the expected mine spectrum on the prior knowledge of the soil wetness (e.g., after a significant rain, the algorithm would expect the spectrum to be closer to $\epsilon_r = 15$, whereas for more typical conditions, a nominal spectrum at (or around) $\epsilon_r = 2.5$ may be used). This would involve a modification of the basic detection algorithm presented here, for which multiple spectra may be considered based on the prior knowledge of soil wetness.

Based on an extensive set of measured data like that in Figs. 1 and 2, and computed data like that in Figs. 1–4, we have observed that the spectral signatures of landmines vary but are relatively robust to variations in the target depths and soil conditions. This is because the energy for a plastic landmine is reverberating principally within the mine, and therefore we still have observed relative robustness of this feature to changing depths and soils. We therefore feel that the spectral characteristics of landmines, particularly plastic landmines, constitute an important classification feature. However, we emphasize that this feature is only useful when placed in the context of other features extracted from the GPR signature. We do not advocate detection and classification of landmines based on the spectral GPR feature alone. As demonstrated in the results presented below, when used as one of several features, the spectral characteristics has proven to provide important classification enhancement using measured field data.

III. ESTIMATION OF EDS FROM GPR DATA

This section describes a methodology to generate the EDS from the GPR response of a target. The aim is to exploit the EDS to improve the detection of weak-scattering plastic landmines and the discrimination between mine target and clutter objects.

The spectral characteristics of weak scattering plastic landmine described in the previous section are found through FDTD modeling. The actual GPR measurements contain ground reflection, background response, and random behavior. Signal processing is therefore necessary to estimate the EDS and obtain the spectral features. We shall first describe the EDS estimation technique. The EDS estimation is based on the periodogram approach [43], and the periodogram is averaged over a spatial window to reduce the estimation variance. To illustrate the consistency of the landmine spectral characteristics and the robustness of the proposed EDS estimation method, the EDS estimator will be applied to two different GPRs. One is wideband pulse-excited radar for a vehicle-mounted system, and the other is a frequency swept handheld system. The former collects data in the time domain and the latter in the frequency domain. The data can be made equivalent if they have the same bandwidth and resolutions in both time and frequency. However, compared to the first GPR, the second has a smaller bandwidth and higher frequency resolution.

In a typical landmine-detection strategy, a prescreener is first applied to indicate the potential locations of mine targets. More sophisticated processing is then followed to affirm if the location has a mine target, to reduce the probability of false alarm (Pfa). The prescreener algorithm has the attribute of high location accuracy and a 100% probability of detection (Pd) with moderate Pfa. Many prescreener algorithms are available; some popular ones include the least mean squares algorithm [44], the constant false alarm rate (CFAR) algorithm [45], principal components analysis (PCA) [46], and correlation detector (CorrDet) [47]. The EDS and the spectral features will be generated on the alarm locations declared by a prescreener, which contain either mines or clutter.

The GPR data at cross-track position x and down-track position y is denoted as $d(x, y, z)$, where z represents depth.

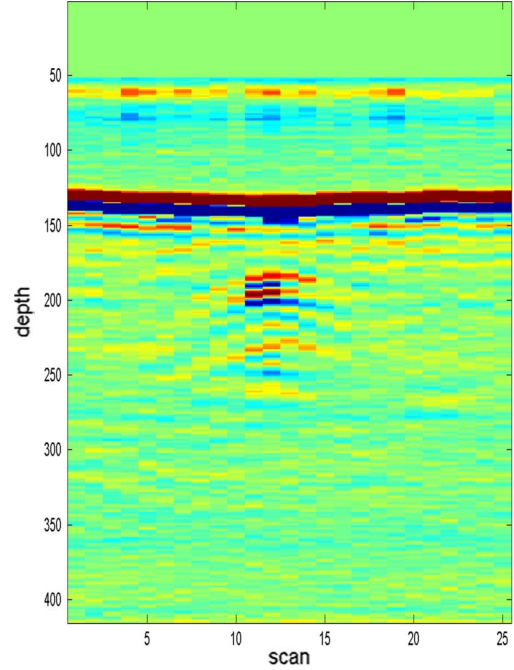


Fig. 5. Section of GPR B-scan that contains a mine target.

A highly simplified data model is shown as follows:

$$d(x, y, z) = g(x, y, z) + s(x, y, z) + w(x, y, z) \quad (1)$$

where $g(x, y, z)$ represents the ground bounce reflection, $s(x, y, z)$ denotes the landmine target or clutter object response, and $w(x, y, z)$ represents the noise. Fig. 5 shows a section of B-scan GPR data that contains a mine in the middle, where the horizontal axis is the down-track and the vertical axis is the depth. The color (gray level) represents the intensity of the GPR signal return. The strongest returned GPR signal is from the ground reflection. For simplicity, we shall call the data collected at surface position (x, y) along the depth a vector sample.

The generation of the EDS contains the following steps: (A) data preprocessing to remove ground reflection, (B) nonlinear smoothing to reduce noise, (C) spectral domain whitening normalization and contrast enhancement, and (D) estimation of spectrum. To evaluate the proposed method in discriminating between mine and clutter objects, a single confidence value will be generated based on the matched filter approach. The details of the different steps are described below.

A. Preprocessing

The purpose of the preprocessing step is to remove the component $g(x, y, z)$. Various techniques to remove the ground effect are available. Here we briefly describe two popular approaches that we will use later.

The first approach is based on range-gating [45]. It estimates the ground level, aligns the data, and processes the data at some distance below the ground only. The depth at which the ground level occurs in a vector sample is estimated as the average in which the maximum and minimum values occur. Data alignment is then applied so that the ground level at each vector sample always occurs at the same place. Only the data at some distance below the ground level are kept for further processing.

The second approach uses the linear-prediction (LP) model to subtract out the ground response [47]. It assumes that the background response at the current vector sample can be formed as a weighted sum of the past few background vector samples. The weighting coefficients (LP coefficients) are different at each sample location and are obtained by the maximum likelihood optimization. The preprocessed vector sample is the difference between the current vector sample and the one based on the background LP model. If the ground level varies significantly, appropriate shifting for ground alignment is needed before computing the LP background estimate for subtraction. The optimum shift in the current vector sample is normally determined in conjunction with the LP coefficient estimation. The LP approach is found to be particularly attractive for handheld GPR systems.

B. Nonlinear Smoothing

Median filtering is applied to each B-scan of the preprocessed data to remove internal GPR noise and any other transient noise. The median filter is 1-D, and the filtering is performed at each depth bin separately.

C. Whitening

The GPR transmit-receive pairs are different at different cross-track positions. The purpose of the whitening step is to remove the internal coupling between the GPR transmit-receive pair and to whiten the background. The internal coupling is relatively constant over different scans for the same GPR transmit-receive pair and is not the same for different GPR transmit-receive pairs. Furthermore, because the GPR operates at a very high frequency in the order of the GHz range, the background data statistics and the internal noise could be slightly different in different GPR transmit-receive pairs. As a result, whitening is performed for each cross-track separately.

After median filtering, the fast Fourier transform (FFT) is applied on each vector sample along depth. Let (x_o, y_o) be the current location of interest. The FFT data before and after the scan at (x_o, y_o) are used to compute the mean $m_D(x_o, k_z)$ and standard deviation $\sigma_D(x_o, k_z)$ of the background for normalization

$$m_D(x_o, k_z) = \frac{1}{2L} \left(\sum_{i=y_o-G-L}^{y_o-G-1} D(x_o, i, k_z) + \sum_{i=y_o+G+1}^{y_o+G+L} D(x_o, i, k_z) \right) \quad (2)$$

$$\sigma_D^2(x_o, k_z) = \frac{1}{2L} \left(\sum_{i=y_o-G-L}^{y_o-G-1} |D(x_o, i, k_z)|^2 + \sum_{i=y_o+G+1}^{y_o+G+L} |D(x_o, i, k_z)|^2 \right) - |m_D(x_o, k_z)|^2 \quad (3)$$

where $D(x_o, y, k_z)$ represents the FFT data at position (x_o, y) , $|(*)|$ is the absolute value of $(*)$, and k_z is the frequency-domain index. Note that $m_D(x_o, k_z)$ is complex and that $\sigma_D^2(x_o, k_z)$ is real. G is the number of guard samples, and L is the number of scans before and after the current location over which to perform averaging. The whitening step is to minimize the effect of soil condition on the EDS of a mine target. When the soil environment is relatively stationary, increasing L can give better background estimate and, hence, better results.

Normalization is then applied to the scans from $y_o - G$ to $y_o + G$, at every frequency bin k_z

$$\tilde{D}(x_o, y, k_z) = \left(\frac{D(x_o, y, k_z) - m_D(x_o, k_z)}{\sigma_D(x_o, k_z)} \right), \quad y = y_o - G, y_o - G + 1, \dots, y_o + G. \quad (4)$$

After whitening, following next is contrast enhancement by removing local mean and semithresholding. The mean and mean-square values of $\tilde{D}(x_o, y, k_z)$ over $y = y_o - G, y_o - G + 1, \dots, y_o + G$ are computed:

$$m_{\tilde{D}}(x_o, k_z) = \frac{1}{2G+1} \sum_{y=y_o-G}^{y_o+G} \tilde{D}(x_o, y, k_z) \quad (5)$$

$$\nu_{\tilde{D}}^2(x_o, k_z) = \frac{1}{2G+1} \sum_{y=y_o-G}^{y_o+G} |\tilde{D}(x_o, y, k_z)|^2. \quad (6)$$

We then subtract out $m_{\tilde{D}}(x_o, k_z)$ from $\tilde{D}(x_o, y, k_z)$, take the absolute value and square, and apply semithresholding at $\nu_{\tilde{D}}^2(x_o, k_z)$, i.e., (7), shown at the bottom of the page. The semithresholding step is to improve the contrast of the EDS estimate and the semithreshold value corresponds to the mean of the background spectra, assuming that the background data is Gaussian distributed. The resultant data $U(x_o, y, k_z)$ is a 2-D matrix with respect to y and k_z . This same procedure is repeated to generate $U(x, y, k_z)$ at other cross-track locations (other x values).

$$U(x_o, y, k_z) = \begin{cases} |\tilde{D}(x_o, y, k_z) - m_{\tilde{D}}(x_o, k_z)|^2, & \text{if } |\tilde{D}(x_o, y, k_z) - m_{\tilde{D}}(x_o, k_z)|^2 \geq \nu_{\tilde{D}}^2(x_o, k_z) \\ 0, & \text{if } |\tilde{D}(x_o, y, k_z) - m_{\tilde{D}}(x_o, k_z)|^2 < \nu_{\tilde{D}}^2(x_o, k_z) \end{cases} \quad (7)$$

D. Spectrum Generation

The spectrum is generated by averaging $U(x, y, k_z)$ over a square window of N samples in cross-track and N samples in down-track

$$P(x_o, y_o, k_z) = \frac{1}{N^2} \sum_{x=x_o-(N-1)/2}^{x_o+(N-1)/2} \sum_{y=y_o-(N-1)/2}^{y_o+(N-1)/2} U(x, y, k_z). \quad (8)$$

Depending on the nature of the data, it is sometimes beneficial to apply median filtering along the cross-track on $U(x, y, k_z)$ before averaging to form the EDS $P(x_o, y_o, k_z)$. The averaging is to reduce the variance in the EDS estimate [43].

E. Spectral Confidence Value

Normally, a spectral feature vector will be produced from the EDS, and it will be used in conjunction with other features obtained in the depth domain to form a detection confidence through an appropriate fusion algorithm. In our study, we generated a confidence value based on the spectral feature vector alone to examine the effectiveness of the EDS in improving mine detection.

There are many ways to obtain a feature or feature vector from the EDS. Described below is just one possible method. Other techniques for generation of the feature vector and confidence value may be more appropriate and give better results, depending on the specific GPR used.

We shall collect $P(x_o, y_o, k_z)$ along k_z to form a spectral feature vector and call it \mathbf{Q} . Depending on the application and the specific GPR, it may be necessary to reduce the number of elements in \mathbf{Q} . A single-feature confidence value is generated using the matched-filter approach. If we use the vector \mathbf{W} to denote the matched filter (after time is reversed), then the spectral correlation feature (SCF) confidence value is

$$SCF = \log(\mathbf{W}^T \mathbf{Q} + 1). \quad (9)$$

The logarithm operation is a nonlinear technique used to compress the dynamic range of the detection confidence value. The matched filter \mathbf{W} is extracted from the weak-scattering plastic mines, either based on training data or the theoretical EDS.

We shall now apply the EDS generation technique to two different radars. The two radars are from two different manufacturers and have different characteristics.

GPR System-1: The prototype GPR system-1 is a vehicle-mounted system for which the GPR sensor is attached to the front of a vehicle [48]. The GPR is a pulse-excited system that measures data in the time domain. The start and stop frequencies of the radar are 200 MHz and 7 GHz, and the sampling rate is 62 GHz. A vector containing 415 data points is collected in each physical location on the ground surface. Because the bandwidth is quite wide, this radar provides a very high resolution in depth. On the other hand, the frequency resolution is low due to the high sampling frequency, and it has a value of

$$\text{FreqResolution} = \frac{62 \times 10^9}{415} = 150 \text{ MHz}. \quad (10)$$

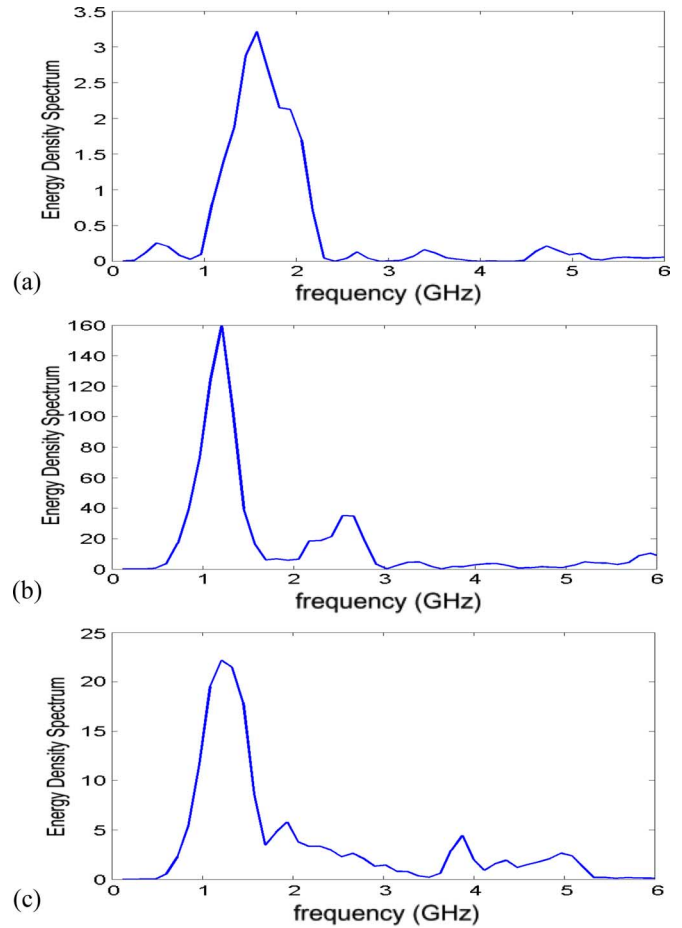


Fig. 6. EDS of three different types of plastic antitank landmines that are known to be difficult to detect. (a) Type-1. (b) Type-2. (c) Type-3.

The GPR data are collected as the vehicle proceeds at every 5 cm down-track and 5 cm cross-track. The CFAR prescreener algorithm [45] processes the data sequentially, and the proposed EDS technique is applied at all declared alarm locations.

The range-gating preprocessing method [45] is used. Pixel-level shifting (not subpixel level) is sufficient to align the data with respect to a global ground level because the radar has a very high sampling frequency. Only the data starting from 25 depth bins below the ground surface is kept for further processing. In nonlinear smoothing, the length of the median filter is 5, which translates to 25 cm because the vector samples are collected at every 5 cm. Zero padding is used to adjust the size of each vector sample to 512 points before applying the FFT. The whitening process uses $G = 6$ guard samples (a distance of 30 cm from the alarm location) and $L = 6$ for background samples. The averaging area in spectrum generation is 25 cm by 25 cm, which corresponds to $N = 5$.

Fig. 6 shows the EDS of three different types of plastic antitank mines: Type-1, Type-2, and Type-3. A Type-1 mine is smaller than the other two types. The Type-1 mine is the same landmine that produces the results in Figs. 1–4. All three mine types are known to have weak scattering and are difficult to detect. An interesting observation is that the EDS from these mines have well-defined spectral peaks, albeit of different amplitudes. The location of the spectral peak in the Type-1 mine is at about 1.6 GHz, whereas that for Type-2 and

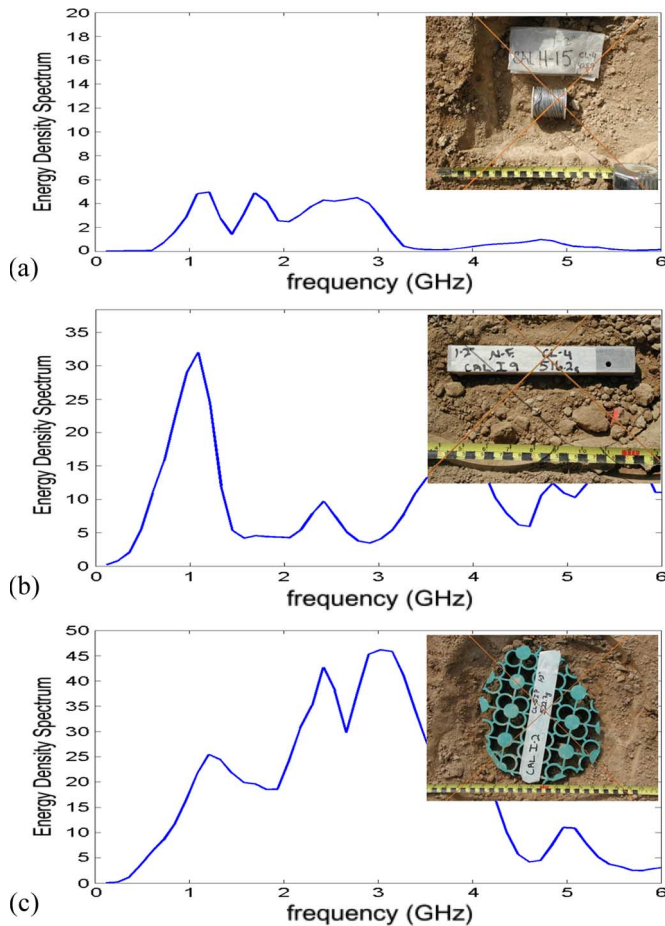


Fig. 7. EDS of three clutter objects. (a) Metal clutter with more than 40 g of metal content. (b) Another metal clutter with more than 40 g of metal content. (c) Piece of irregular plastic.

Type-3 mines is about 1.2 GHz. The spectral peak frequency locations are consistent with the first spectral peak location in the theoretical study given in Section II (see Figs. 3 and 4). The spectral peaks at higher frequencies are not as apparent as in the theoretical study. There are two explanations. First, frequency normalization, or whitening, is performed when the EDS (Section III-C) is generated. The background response has larger variations in high frequencies so that the peaks at high-frequency locations are suppressed after whitening. This is not considered to be a disadvantage from the proposed EDS estimation method because the high-frequency peaks are less reliable and therefore have relatively small impact in improving performance. Second, the results shown in Figs. 1–4 are obtained when the radar signal is impinging perpendicular to the target at the center. The spectrum generation step in Section III-D performs averaging over a cross-track by down-track window to reduce estimation variance, whereas the scans within the spatial window have different radar incident angles with respect to the target. As a result, averaging could reduce the high-frequency peaks.

Fig. 7 shows the EDS of three different clutter objects. Clutter objects 1 and 2 have a metal content larger than 40 g but have different shapes, and object 3 is a piece of irregular plastic. These clutter objects have different shapes and compositions than a mine target; they are also shown in Fig. 7. The clutter

objects all have a strong GPR energy return. Interestingly though, their spectra have shapes quite different from those of the mines, especially in clutter objects 1 and 3. As a result, it is expected that the EDS will be useful in increasing the detection of some weak-scattering plastic mines, and at the same time providing discrimination ability between mine and clutter objects. It should be noted that there is a limit to which the EDS can be provided in the discrimination between landmine and clutter objects.

The shape of the EDS is resulted from a target's height, shape, and composition. It is, therefore, difficult to distinguish a mine from a clutter object using the EDS if the clutter object has similar height, shape, or composition as a landmine. Based on our experiments and data collection, a soda can could create an EDS similar to that of a mine, and the EDS of irregular plastics and pieces of wood can be distinguished from that of a mine more easily. How well we can distinguish a mine with a clutter could largely depend on the probability of detection. We may be able to distinguish a certain type of mine with clutter relatively well by setting a high detection threshold, but the probability of detection over a variety of mine types could be very low because the EDS from different mines have variations, and the EDS of the same mine type could also vary under different orientations and environmental conditions. On the other hand, most mine fields have limited types of landmine targets. If some prior knowledge is available about the few types of landmines to detect, an algorithm using EDS can be "tuned" to detect these certain types of targets to improve the discrimination between landmines and clutter objects.

Fig. 8 gives the EDS of clutter objects with different levels of metal content. The first one has less than 3 g of metal, the second one has 3–10 g, and the third one has more than 40 g. The images of the three clutter objects are also shown in the figure. When the levels of clutter increase, the magnitude of the EDS increases. This is because the area under the EDS curve corresponds to the total energy return of the clutter objects. However, increasing the clutter level will not give spectral characteristics similar to a landmine.

The consistency of the landmine spectral characteristics and the robustness of the proposed EDS generation technique are illustrated in Fig. 9, which shows the EDS of Type-1 landmines derived from the data collected at three different sites. Site-1 has dry soil in an arid climate, and the soil types of Site-2 and Site-3 contain both dirt and gravel in a temperate climate. The three sites are geographically separated in different parts of the United States. In each site, three EDS are shown that correspond to the Type-1 mines at different depths. The nine EDS are from different Type-1 mines. Within each site, the EDS are very similar and insensitive to the depth of the mine targets as anticipated from the theory. Among the different sites, the EDS are quite consistent, although the three sites have significant different soil characteristics.

We have computed the correlation coefficients for spectra in Fig. 9, and they are shown in Table I. Element (i, j) in the matrix is the correlation coefficient between EDS i and EDS j , where the EDS are numbered in their order of appearance in Fig. 9. To take into account the slight shift in the EDS peak location, the correlation coefficients within a shifting range of ± 480 MHz in EDS j is computed, and the maximum of them is

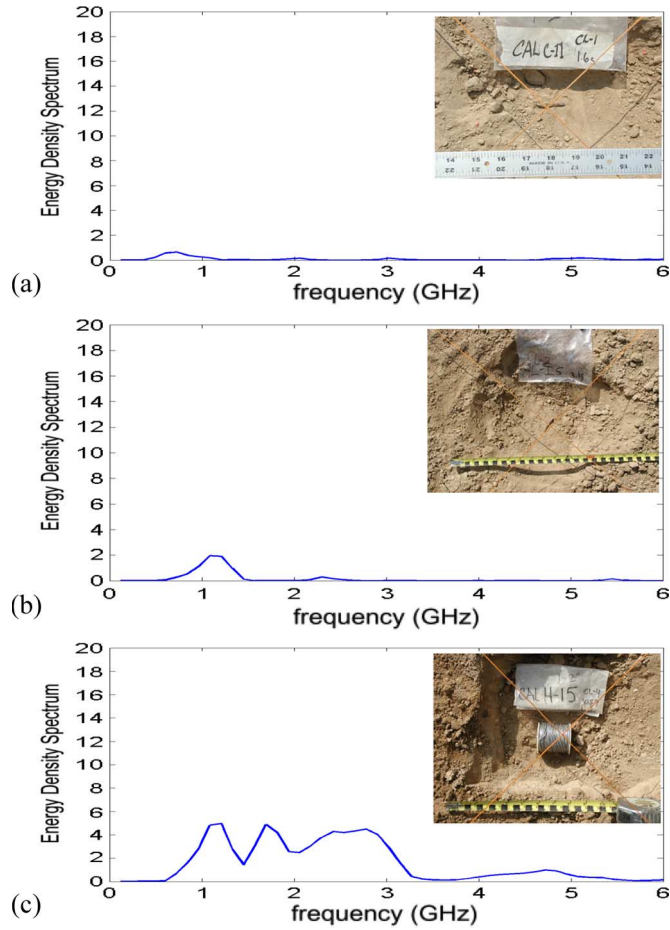


Fig. 8. EDS of clutter objects by varying amounts of clutter. (a) Metal clutter with less than 3 g of metal. (b) Metal clutter with 3–10 g of metal. (c) Metal clutter with more than 40 g of metal.

the value given in the matrix. The matrix has unity in the main diagonal as expected. The smallest value in the matrix is 0.84, which is considered close to 1. Because of the relatively small number of vector samples available (only 25) in estimating the EDS, the spectral peak locations in EDS tend to drift slightly, even for the same type of landmines as depicted in Fig. 9(c). The EDS may also vary a little due to different soil conditions (see Fig. 9) and the varieties of plastic mine types. To take these factors into account, and to increase the robustness of the proposed technique, the spectral energies at different frequency bands will be used to form the spectral feature vector. The size of each frequency band is set to 600 MHz. Hence, there will be ten spectral features over the frequency range up to 6 GHz. Note that the FFT size is 512, and the sampling frequency is 62 GHz. The frequency bin size is therefore $62/512 = 120$ MHz, and each frequency band covers $600/120 = 5$ frequency samples. The frequency bands are decomposed by using a cosine square window.

To be more specific, the j th spectral feature is generated by

$$Q(x_o, y_o, j) = \sum_{i=-(M-1)/2}^{(M-1)/2} P(x_o, y_o, Bj + i) \cos^2\left(\frac{\pi}{M}i\right) \quad (11)$$

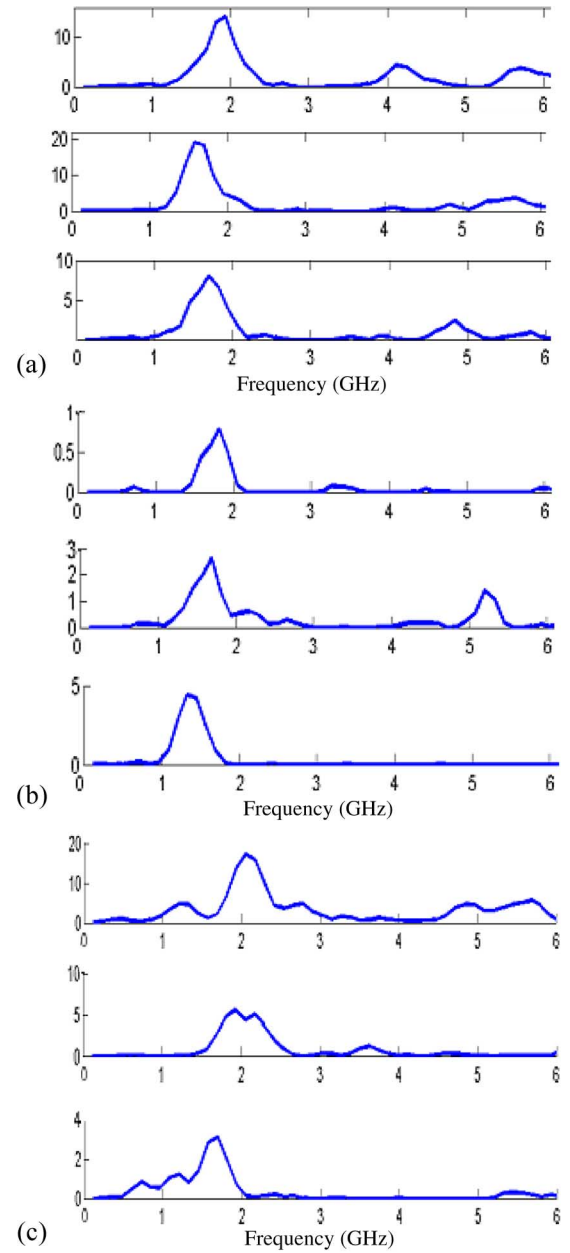


Fig. 9. EDS of a Type-1 mine at three different test sites that are geographically separated with different soil conditions. (a) Mine depth at Site 1 is 5.1 cm (above), 7.6 cm (middle), and 12.7 cm (bottom). (b) Mine depth at Site 2 is 5.1 cm (above), 7.6 cm (middle), and 10.2 cm (bottom). (c) Mine depth at Site 3 is 0 cm (above), 5.1 cm (middle), and 10.2 cm (bottom).

where B is the frequency subband size that is set to 5, and M is the window width and is equal to $M = 2B - 1$. There is 50% overlap between the two adjacent subbands, and j takes values from 1 to 10. The collection of $Q(x_o, y_o, j)$ with respect to j forms the 10-element feature vector \mathbf{Q} .

To generate a test statistic, a matched filter is designed to match with the \mathbf{Q} values computed in (11). The matched filter is derived by computing the average of 15 EDS measured at a test site for a Type-1 landmine buried at depths between 5.1 to 12.7 cm, and the filter coefficients are rounded to a single digit. The EDS were normalized with the maximum value equal to one before subbanding and averaging. The values above 3 GHz were set to zero in the subband-averaged EDS because

TABLE I
MAXIMUM VALUES OF THE CROSS-CORRELATION COEFFICIENTS
FROM THE ENERGY DENSITY SPECTRA IN FIG. 9

Corr. Coef	1	2	3	4	5	6	7	8	9
1	1	0.95	0.91	0.88	0.85	0.91	0.84	0.89	0.90
2		1	0.94	0.94	0.88	0.94	0.88	0.92	0.91
3			1	0.95	0.90	0.95	0.91	0.94	0.91
4				1	0.87	0.98	0.86	0.91	0.90
5					1	0.88	0.94	0.85	0.87
6						1	0.86	0.91	0.92
7							1	0.85	0.86
8								1	0.88
9									1

the EDS vary significantly above 3 GHz, and the EDS values above 3 GHz is not reliable. The resulting matched filter is

$$\mathbf{W} = [0.2, 0.4, 1, 0.4, 0.2, 0, 0, 0, 0]^T. \quad (12)$$

It has the largest value at 1.5 GHz and has nonzero values over the frequency range up to 3 GHz.

The design of \mathbf{W} should be based on physical derivations or an extensive data collection. A better design of the matched filter could yield a better result. We would like to point out that with sufficient amount of data for different weak-scattering landmines, clustering technique could be used to obtain several matched filters, instead of one, to improve performance.

To design a better classifier that uses EDS features, we shall look into the resampling techniques such as jackknife and bootstrap [49]–[51], to deduce the knowledge about the statistical distributions of landmine EDS for classification. We also plan to examine the use of some robust classification techniques [52], [53] to improve the classifier design.

GPR System-2: The GPR System-2 is a handheld based system where the GPR sensor is attached to the tip of a handheld unit. It is a frequency-swept radar, and the bandwidth is 1.4 GHz only, from 1.1 to 2.5 GHz. However, it has a much higher frequency resolution of 20 MHz, which is about 7.5 times larger than the vehicle-mounted radar. The spatial sampling spacing is also denser, in which about 160 samples were collected in 50 cm.

Initially, the operator sweeps the detector back and forth in the cross-track direction while moving forward to collect the GPR data. The PCA-based prescreening algorithm [46] is used to generate initial declarations of a potential mine target. Once a potential mine target location is identified, the detector will go to the discrimination mode where the operator will stand still and sweep the detector back and forth over the alarm center to collect more data to ascertain if this location contains a landmine target. The EDS technique will be applied to the discrimination mode data in each sweep separately.

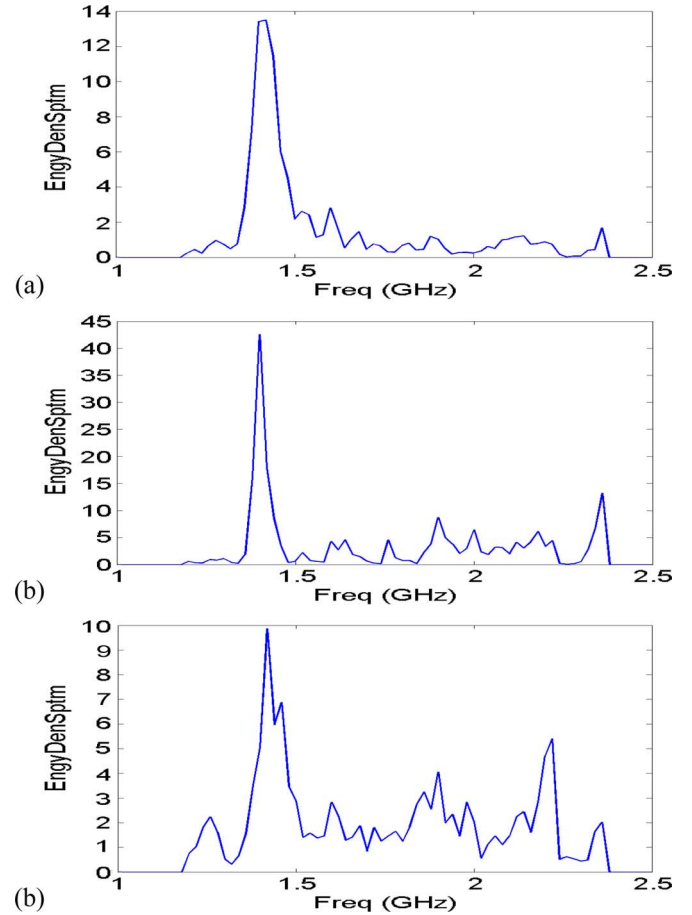


Fig. 10. EDS of the same three plastic antitank mines as in Fig. 6, where the data were collected by the handheld radar that had a different bandwidth and sampling frequency.

In the EDS generation, preprocessing uses the LP technique [47] to remove the ground reflection. The nonlinear smoothing uses a size 15 median filter. The size is larger than that used in the previous GPR because of denser spatial sampling in the handheld system. The parameters G and L in the whitening process have different values for each suspected object, where G is determined to be the number of samples having energy values larger than 15% of the maximum energy of the preprocessed and nonlinearly smoothed data in a single sweep, and L corresponds to the number of samples that have energies below 15%. The 60% and 15% figures were selected based on experimentation to obtain the best performance from GPR System-2. Because the handheld system collects data in 1-D sweeps, the averaging in EDS is over the cross-track direction x only, and the averaging size N within a sweep corresponds to the samples with energies above 60% of the maximum energy in the sweep. Because the radar has a much denser spatial sampling and a finer spectral resolution, a more accurate estimate of the EDS can be obtained. As a result, we use all 70 frequency points to form the feature vector, and no subbanding was applied. The matched filter to be multiplied with the spectral feature vector as indicated in (9) is generated from the training data.

Using the data from the handheld GPR, Fig. 10 depicts the EDS of the same three mines corresponding to those in Fig. 6. Unlike the pulsed radar, the handheld GPR (frequency-swept GPR) sends out sinusoids of frequencies separated by

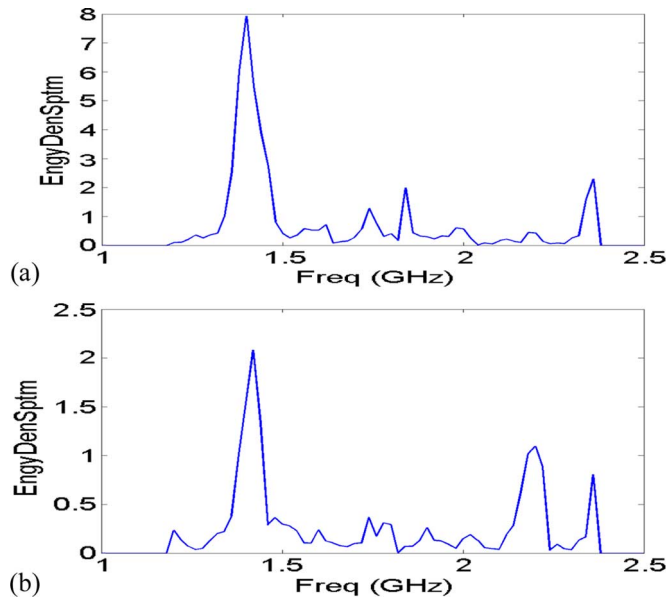


Fig. 11. EDS of two antipersonnel mines, data from the handheld radar having smaller bandwidth but a higher frequency resolution.

20 MHz and measures the signal strengths returned back. Hence, the EDS of the handheld GPR has a frequency resolution of 20 MHz, giving a finer frequency resolution that is 7.5 times higher than that of the pulsed radar. As a result, the EDS produced by the frequency-swept radar (Fig. 10) have much sharper spectral peaks than the EDS from the pulsed radar (Fig. 6). The difference in the design of the two radars—pulsed versus frequency swept—contributes to the difference in peak size of the EDS from the two radars. It should be noted that the total bandwidth of the frequency-swept GPR is only 1.4 GHz, so that its resolution in depth is much smaller than the pulsed GPR that has a bandwidth of 7 GHz. The consequence is that the mine target and the ground have very few depth pixels separation in the frequency-swept radar, and LP technique is needed to remove the ground bounce effect instead of using range gating as in the pulsed radar. We observe that the EDS in Figs. 6 and 10 are quite consistent with each other in the spectral peak location. The consistency to some extent corroborates the fact that the EDS features result from the physical characteristics of mine targets, and they are relatively insensitive to the GPR used.

The EDS characteristics occur not only in the plastic antitank mines, they also appear in low-metal antipersonnel mines that are much smaller than the antitank mines. Fig. 11 gives the EDS plots of two weak-scattering plastic antipersonnel mines derived from the handheld GPR data. Interestingly enough, their spectral characteristics are very similar to those for the three plastic antitank mines shown in Fig. 10.

IV. EXPERIMENTAL RESULTS

A number of experiments were performed to corroborate the effectiveness of the spectral features to improve the detection of landmines. The first experiment uses the data collected from the wide bandwidth vehicle-mounted GPR. The second experiment applies to the data collected from the smaller-bandwidth handheld-based GPR. The third experiment examines the fu-

sion performance of the spectral feature confidence value and detection confidence produced from the time-domain geometric features extracted from the GPR signature.

The three experiments use different datasets. They were obtained from different test sites that are geographically separated and have different soil properties and conditions. The first dataset was collected in October 2002 and January 2003, the second in October 2004, and the third in February 2004. The data were collected over lanes that contained both landmine targets and clutter objects. The clutter objects could be pieces of metal, pieces of wood, plastic caps, and many variations of them. The feed-forward ordered weighted averaging (FOWA) algorithm [45] that is based on the geometric features generated from the time-domain GPR data were used as a baseline reference for comparison.

A. Experiment 1

The first experiment used the data collected in a desert region with a dry soil condition. The data were collected twice, in October 2002 and January 2003. The mine lane was 50 m long and 3 m wide and contained 11 plastic antitank plastic landmines from three different types buried at depths of either 7.6 or 12.7 cm. Seven of those 11 mines were known to have weak-scattering GPR signal and were therefore difficult to detect in a typical time-domain approach. Table II(a) and (b) show the results of the two data collections in terms of number of mines detected and the corresponding false alarm rate (FAR). The FAR was computed by dividing the number of false alarms by the area of the lane. The SCF confidence value did not have any geometric information because the EDS was generated by averaging over a square window of 25 cm cross-track and 25 cm down-track. It was therefore beneficial to divide the SCF confidence given in (9) by the fixed compactness to form the overall detection confidence. The fixed compactness [45] is the radius of a disk centered at the alarm location that contains 45% of the total GPR energy projected along several depth segments. The results from using $1/\text{Compactness}$ as the confidence and the FOWA algorithm output are also given for comparison. FOWA [45] computed the geometric features such as compactness, solidity, and eccentricity from the time-domain GPR-whitened signal energies at several depth segments, found the ordered weighted averages (OWA) of them, and used a decision network to form the confidence for landmine detection. The FOWA score is the average of three independent trainings that started with different initial seeds. Although the average in the FOWA scores was taken, the variations of the individual FOWA scores were very small. The testing data were not included in the training. The receiver-operating characteristic (ROC) curves corresponding to the results given in Table II is given in Figs. 12 and 13.

The results in Table II(a) and (b) are quite consistent. In particular, the detection results from the spectral confidence divided by the compactness are better than those from $1/\text{Compactness}$, as well as the FOWA scores. For the first dataset, FOWA is not able to reach 100%, and in the second dataset, the reduction in FAR over FOWA is 35% at 100% Pd. It is evident that the spectral feature is able to reduce the number of false alarms produced by clutter objects and improve the detection of weak mines.

TABLE II
(a) DETECTION PERFORMANCE OF THE OCTOBER 2002
COLLECTION OF DATASET 1 IN EXPERIMENT 1;
(b) DETECTION PERFORMANCE OF THE JANUARY 2003
COLLECTION OF DATASET 1 IN EXPERIMENT 1

# Mines Detected (Tot=11)	FAR (/m ²)		
	SCF/Compactness	1/Compactness	FOWA
11	0.012	0.023	N/A
10	0.011	0.016	0.021
9	0.0045	0.015	0.021
8	0.0045	0.015	0.021
7	0.0045	0.015	0.014
6	0.0045	0.011	0.012
5	0.0034	0.010	0.0034
4	0.0023	0.0068	0.0026
3	0.0023	0.0056	0.0023
2	0.0011	0.0034	0.0015
1	0.0	0.0	0.0

(a)

# Mines Detected (Tot=11)	FAR (/m ²)		
	SCF/Compactness	1/Compactness	FOWA
11	0.017	0.033	0.023
10	0.0056	0.015	0.022
9	0.0056	0.011	0.012
8	0.0045	0.011	0.012
7	0.0045	0.011	0.0086
6	0.0034	0.0079	0.0083
5	0.0011	0.0067	0.0041
4	0.0	0.0034	0.0034
3	0.0	0.0023	0.0008
2	0.0	0.0011	0.0008
1	0.0	0.0	0.0

(b)

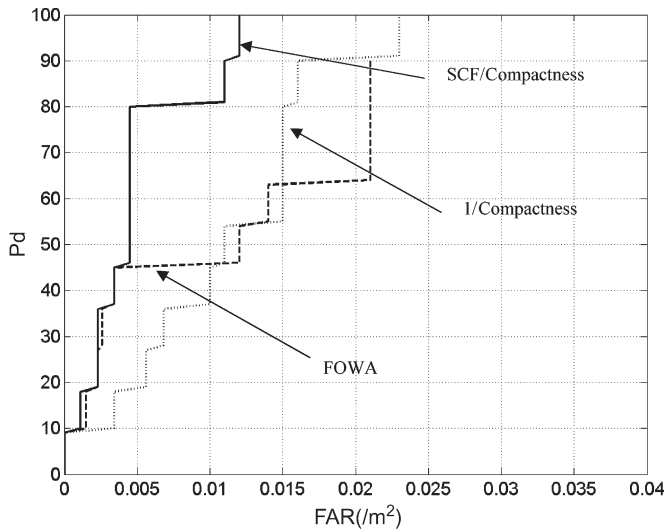


Fig. 12. ROC curves that correspond to the results in Table II(a).

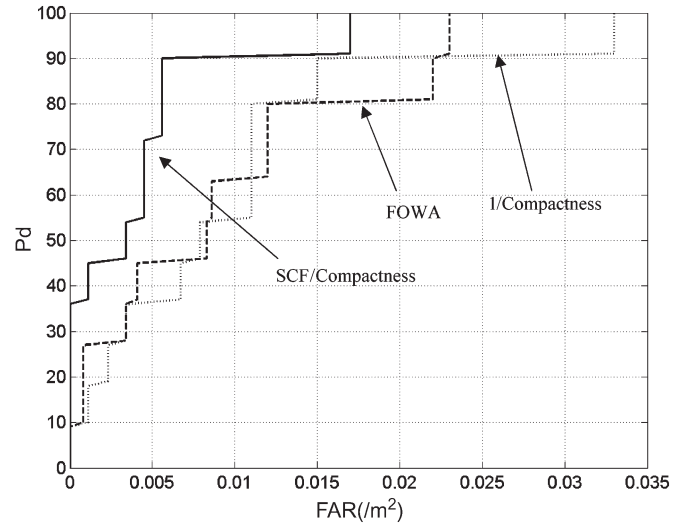


Fig. 13. ROC curves that correspond to the results in Table II(b).

B. Experiment 2

This experiment used the second dataset collected in October 2004 using the handheld frequency-swept GPR, which has a smaller bandwidth compared to the vehicle-mounted GPR. The data were acquired from another test site different from that in Experiment 1. Unlike the first experiment, in which the mine and clutter objects were buried randomly in a lane, the mine and clutter objects are at the centers of 1-m² cells, and they are buried at depths ranging from 0.625 to 10.16 cm. There were 44 mine targets, 158 clutters, and 23 empty cells. There were 12 types of landmines, and the landmine targets had good mix and variations of antitank, antipersonnel, plastic, and metal mines. Regarding clutter, both metal and nonmetal clutter objects were present. Metal clutter had metal content ranging from less than 3 g to more than 40 g. Nonmetal clutter included plastic, stones, and pieces of wood in regular and irregular shapes.

When the data were collected, the GPR sensor head swept the suspected alarm location six times across, and two times in the perpendicular direction. The (6,2) sweeps were decided based on a balance between performance and the time needed to collect the sweeps. Among these eight sweeps, two or three were selected based on the largest energy return, where two sweeps was for antipersonnel mines and three sweeps is for antitank mines. The EDS were then generated for the selected sweeps using the procedure as described in Section III. The EDS were then averaged to form a single EDS from which to compute the SCF confidence value according to (9). There were three matched filters, each corresponding to a particular set of weak-scattering antipersonnel mines. The maximum of the three matched filter outputs is the SCF confidence.

Compactness was also used here to improve the confidence value because the SCF did not contain size information. The sweeps in this case are only 1-D, and the compactness is generated based on the 1-D sweep, which is defined as the square root of the successive number of samples whose projected energies along depth were bigger than 60% of the largest energy value in this sweep. The confidence value for scoring was the SCF divided by this compactness value. This radar has a high-frequency resolution, and we found that some metal clutter

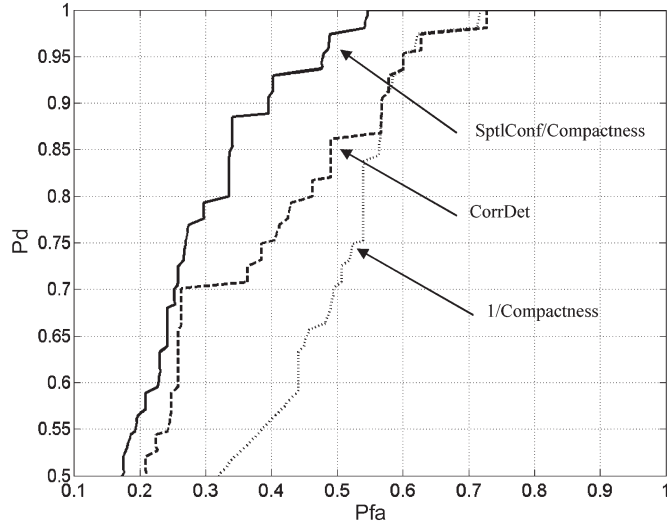


Fig. 14. ROC curves from SCF/Compactness, 1/Compactness, and CorrDet.

objects have significant energies at the frequency band from 1.1 to 1.3 GHz relative to the mine target frequency band ranged from 1.3 to 1.5 GHz. Hence, we decreased the confidence value by a factor of two if the ratio of the energies in the bands between (1.1–1.3 GHz) and (1.3–1.5 GHz) was larger than 0.4. This procedure was determined experimentally, and a better formulation was needed to make better use of the EDS to capture the difference in mine and clutter characteristics.

Fig. 14 shows the ROC curve comparisons of this radar. The FOWA algorithm is not applicable for this dataset because it contains only 1-D sweeps, and the features in FOWA assumes 2-D energy maps at different depth segments. Hence, the result from CorrDet [47] was used for comparison, where the single confidence value was obtained by taking the maximum of the CorrDet algorithm output in the selected sweeps for each target object. The proposed spectral feature was able to decrease the Pfa by 35% at 90% Pd, which is a very significant improvement. Also shown is the ROC curve computed from 1/Compactness only. The result from 1/Compactness is quite worse, which confirms that the source of the improvement is the spectral feature.

There is one type of antipersonnel landmine that was relatively difficult to detect. Its low detectability was because this type of landmine has a very low energy return, and hence the signal-to-background noise ratio was very low, and the EDS generated was not able to improve its detectability significantly. Regarding clutter, we find that a slightly compressed soder can produce very similar spectral characteristics as a landmine because its shape is very close to some small antipersonnel landmines.

C. Experiment 3

The third experiment examined the advantage of fusing the detection confidences from spectral features and geometric features extracted from GPR signature to improve performance. The dataset was collected by the vehicle-mounted mine detection system in February 2004 but at a different site from the previous two experiments. Here, we used the simple geometric mean fusion with an offset as shown below:

$$\text{Fused Result} = \text{FOWA} \bullet (\text{SCF/Compactness} + 0.5) \quad (13)$$

TABLE III
FUSION PERFORMANCE IN TWO MINE LANES AND ONE CLUTTER LANE

Pd (%)	Number of False Alarms			
	FOWA		Fusion	
	All Lanes	Clutter Lane	All Lanes	Clutter Lane
99 – 100	65	9	79	12
98 – 99	54	7	12	1
96 – 97	52	6	9	1
95 – 96	32	2	7	1
94 – 95	29	2	7	1
93 – 94	22	2	7	1

where FOWA denotes the confidence value from the FOWA algorithm, which was obtained from the geometric features only. The constant 0.5 was determined based on the dynamic range of the FOWA and SCF/Compactness confidence values. It had the effect of setting the fused confidence to be half of that from FOWA when the SCF/Compactness value was very low. This would happen for some wooden box mines in which their spectral characteristics did not show a spectral peak around 1–2 GHz.

To examine the ability of using the spectral feature to reduce the number of false alarms due to clutter objects, we scored two mine lanes and one clutter lane together. There were a total of 80 mines from the two mine lanes. The clutter lane had emplaced clutter objects. The two mine lanes had a total area of about 850 m², and the clutter lane had an area of 300 m². Table III shows the number of false alarms resulting from FOWA and after fusing it with the SCF/Compactness using (13). The first column is the Pd, the second column is the total number of false alarms for FOWA, the third column is the false alarm count in the clutter lane only, and the fourth and fifth columns are the corresponding false alarm counts after fusion. It can be seen that, except at 100% Pd, fusing FOWA with SCF spectral confidence reduced the number of false alarms significantly, particularly when the Pd was around 96% to 99%.

V. CONCLUSION

This paper investigated the spectral characteristics of a target obtained by GPR measurements to improve landmine detection and clutter discrimination. We began with the theoretical study of the EDS of some weak-scattering plastic landmines through FDTD modeling and derived an estimation procedure that generated the EDS at an alarm location using GPR data measurements. Both theory and experimental study revealed that the EDS of some weak-scattering plastic landmines had distinct characteristics, which can be exploited for their discrimination with clutter objects to improve their detection. The consistency of the landmine spectral characteristics were confirmed by the data collected at several geographically separated test sites having different soil conditions and by the data produced from two completely different radar systems. The experimental results corroborated the effectiveness of the spectral features in improving landmine/clutter discrimination and the robustness of the EDS estimation method.

The EDS was able to discriminate the clutter objects that had different size, geometry and composition with a landmine target. In practice, there may be some clutter objects that have similar geometry and composition characteristics as a landmine. As a result, the proposed technique will be more appropriate to be used as features that will be fused with other algorithm outputs. Our recent work [54] indicates that by fusing the spectral features from EDS with the time-domain GPR features, or metal detector features, better performance can be achieved.

ACKNOWLEDGMENT

The authors would like to thank Pete Howard, Denis Reidy, Peter Ngan, Sean Burke, and Richard Weaver for their support and sponsorship. The authors would also like to thank the many reviewers and the Associate Editor of the journal for providing valuable comments and suggestions that have resulted in improving the technical content and presentation of the manuscript. We are particularly grateful to the reviewer who suggested the use of a measure to determine the similarity of the EDS for comparison and the approaches and direction to design a better classifier.

REFERENCES

- [1] R. Seigel, "Land mine detection," *IEEE Instrum. Meas. Mag.*, vol. 5, no. 4, pp. 22–28, Dec. 2002.
- [2] R. B. Cosgrove, P. Milanfar, and J. Kositsky, "Trained detection of buried mines in SAR images via the deflection-optimal criterion," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2569–2575, Nov. 2004.
- [3] Y. Zhang, X. Liao, and L. Carin, "Detection of buried targets via active selection of labeled data: Application to sensing subsurface UXO," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2535–2543, Nov. 2004.
- [4] A. Muscio and M. A. Corticelli, "Land mine detection by infrared thermography: Reduction of size and duration of the experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1955–1964, Sep. 2004.
- [5] G. A. Clark, S. K. Sengupta, W. D. Aimonetti, F. Roeske, and J. G. Donnetti, "Multispectral image feature selection for land mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 304–311, Jan. 2000.
- [6] C. T. Schroder and W. R. Scott, Jr., "A finite-difference model to study the elastic-wave interactions with buried land mines," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1505–1512, Jul. 2000.
- [7] L. Carin, R. Kapoor, and C. E. Baum, "Polarimetric SAR imaging of buried landmines," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 6, pp. 1985–1988, Nov. 1998.
- [8] C. Bruschini, "On the low-frequency EMI response of coincident loops over a conductive and permeable soil and corresponding background reduction schemes," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1706–1719, Aug. 2004.
- [9] I. J. Won and H. Huang, "Automated identification of buried landmines using normalized electromagnetic induction spectroscopy," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 640–651, Mar. 2003.
- [10] L. Collins, P. Gao, D. Schofield, J. P. Moulton, L. C. Makowsky, D. M. Reidy, and R. C. Weaver, "A statistical approach to landmine detection using broadband electromagnetic induction data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 950–962, Apr. 2002.
- [11] T. H. Bell, J. B. Barrow, and J. T. Miller, "Subsurface discrimination using electromagnetic induction sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1286–1293, Jun. 2001.
- [12] L. Carin, Y. Haitao, Y. Dalichaouch, A. R. Perry, P. V. Czipott, and C. E. Baum, "On the wideband EMI response of a rotationally symmetric permeable and conducting target," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1206–1213, Jun. 2001.
- [13] I. J. Won, D. A. Keiswetter, and T. H. Bell, "Electromagnetic induction spectroscopy for clearing landmines," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 703–709, Apr. 2001.
- [14] L. Collins, G. Ping, and L. Carin, "An improved Bayesian decision theoretic approach for land mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 811–819, Mar. 1999.
- [15] L. He, S. Ji, W. R. Scott, and L. Carin, "Adaptive multimodality sensing of landmines," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1756–1774, Jun. 2007.
- [16] P. D. Gader, J. M. Keller, O. Sjahputera, and T. Wang, "Frequency subband processing and feature analysis of forward-looking ground-penetrating radar signals for land-mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 718–729, Mar. 2007.
- [17] S. Lambot, O. Lopera, N. Milisavljevic, and E. C. Slob, "Filtering soil surface and antenna effects from GPR data to enhance landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 707–717, Mar. 2007.
- [18] Y. Liao, L. W. Nolte, and L. M. Collins, "Decision fusion of ground-penetrating radar and metal detector algorithms—A robust approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 398–409, Feb. 2007.
- [19] W.-H. Lee, P. D. Gader, and J. N. Wilson, "Optimizing the area under a receiver operating characteristic curve with application to landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 389–397, Feb. 2007.
- [20] R. Firoozabadi, E. L. Miller, C. M. Rappaport, and A. W. Morgenthaler, "Subsurface sensing of buried objects under a randomly rough surface using scattered electromagnetic field data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 104–117, Jan. 2007.
- [21] T. G. Savelyev, L. van Kempen, H. Sahli, J. Sachs, and M. Sato, "Investigation of time-frequency features for GPR landmine discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 118–129, Jan. 2007.
- [22] D. Potin, E. Duflos, and P. Vanheeghe, "Landmines ground-penetrating radar signal enhancement by digital filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2393–2406, Sep. 2006.
- [23] J. Song, Q. H. Liu, P. Torrione, and L. Collins, "Two-dimensional and three-dimensional NUFFT migration method for landmine detection using ground-penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1462–1469, Jun. 2006.
- [24] D. Potin, P. Vanheeghe, E. Duflos, and M. Davy, "An abrupt change detection algorithm for buried landmines localization," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 260–272, Feb. 2006.
- [25] Q. Zhu and L. M. Collins, "Application of feature extraction methods for landmine detection using the Wichmann/Niitek ground-penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 81–85, Jan. 2005.
- [26] P. D. Gader, W. Lee, and J. N. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.
- [27] X. Millard and Q. H. Liu, "Simulation of near-surface detection of objects in layered media by the BCGS-FFT method," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 2, pp. 327–334, Feb. 2004.
- [28] B. Sai and L. P. Ligthart, "GPR phase-based techniques for profiling rough surfaces and detecting small, low-contrast landmines under flat ground," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 2, pp. 318–326, Feb. 2004.
- [29] K. C. Ho, L. M. Collins, L. Huttel, and P. D. Gader, "Discrimination mode processing for EMI and GPR sensors for hand-held land mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 249–263, Jan. 2004.
- [30] J. M. Stiles, A. V. Apte, and B. Beh, "A group-theoretic analysis of symmetric target scattering with application to landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1802–1814, Aug. 2002.
- [31] X. Xu, E. Miller, C. M. Rappaport, and G. D. Sower, "Statistical method to detect subsurface objects using array ground-penetrating radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 963–976, Apr. 2002.
- [32] X. Xu, E. Miller, and C. M. Rappaport, "Minimum entropy regularization in frequency-wavenumber migration to localize subsurface objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 8, pp. 1804–1812, Aug. 2003.
- [33] A. Van Der Merwe and I. J. Gupta, "A novel signal processing technique for clutter reduction in GPR measurements of small, shallow land mines," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2627–2637, Nov. 2000.
- [34] C. Chen, S. Nag, W. D. Burnside, J. I. Halman, K. A. Shubert, and L. Peters, Jr., "A standoff, focused-beam land mine radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 507–514, Jan. 2000.
- [35] K. O'Neill, "Radar sensing of thin surface layers and near-surface buried objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 480–495, Jan. 2000.
- [36] H. Brunzell, "Detection of shallowly buried objects using impulse radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 875–886, Mar. 1999.
- [37] T. P. Montoya and G. S. Smith, "Land mine detection using a ground-penetrating radar based on resistively loaded Vee dipoles," *IEEE Trans. Antennas Propag.*, vol. 47, no. 12, pp. 1795–1806, Dec. 1999.

- [38] C.-C. Chen and L. Peters, Jr., "Buried unexploded ordnance identification via complex natural resonances," *IEEE Trans. Antennas Propag.*, vol. 45, no. 11, pp. 1645–1654, Nov. 1997.
- [39] C.-C. Chen and L. Peters, Jr., "Ground penetration radar target classification via complex natural resonances," in *Proc. IEEE Antennas Propag. Int. Symp.*, Jun. 1995, vol. 3, pp. 1586–1589.
- [40] S. Vitebskiy and L. Carin, "Resonances of perfectly conducting wires and bodies of revolution buried in a lossy dispersive half-space," *IEEE Trans. Antennas Propag.*, vol. 44, no. 12, pp. 1575–1583, Dec. 1996.
- [41] K. Kim and W. R. Scott, "Design and realization of a discretely loaded resistive vee dipole for ground-penetrating radars," *Radio Sci.*, vol. 39, no. 4, pp. RS4S03.1–RS4S03.9, Jul. 2004.
- [42] T. J. Gander, *Jane's Ammunition Handbook*. Janes, 1998.
- [43] J. G. Proakis, C. M. Rader, F. Ling, C. L. Nikias, M. Moonen, and I. K. Proudler, *Algorithms for Statistical Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [44] P. Torriane, L. Collins, F. Clodfelter, S. Frasier, and I. Starnes, "Application of the LMS algorithm to anomaly detection using the Wichmann/Niitek ground penetrating radar," in *Proc. SPIE—Detection Remediation Technologies Mines Minelike Targets VIII*, Orlando, FL, Apr. 2003, pp. 1127–1136.
- [45] P. D. Gader, R. Grandhi, W. Lee, J. Wilson, and K. C. Ho, "Feature analysis for the NIITEK ground-penetrating radar using order-weighted averaging operators for landmine detection," in *Proc. SPIE—Detection Remediation Technologies Mines Minelike Targets IX*, Orlando, FL, Apr. 2004, pp. 953–962.
- [46] S. Yu, R. K. Mehra, and T. R. Witten, "Automatic mine detection based on ground penetrating radar," in *Proc. SPIE—Detection Remediation Technologies Mines Minelike Targets IV*, Orlando, FL, Apr. 1999, pp. 961–972.
- [47] K. C. Ho and P. D. Gader, "A linear prediction land mine detection algorithm for hand held ground penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1374–1384, Jun. 2002.
- [48] L. M. Collins, P. Torriane, V. Munshi, C. S. Throckmorton, Q. Zhu, F. Clodfelter, and S. Frasier, "Algorithms for landmine detection using NIITEK ground penetrating radar," in *Proc. SPIE—Detection Remediation Technologies Mines Minelike Targets VII*, Orlando, FL, Apr. 2002, pp. 709–718.
- [49] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*. Bristol, PA: SIAM, 1982.
- [50] H. T. Ong and A. M. Zoubir, "The bootstrapped matched filter and its accuracy," *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 11–13, Jan. 2000.
- [51] H. T. Ong and A. M. Zoubir, "Bootstrap-based detection of signals with unknown parameters in unspecified correlated interference," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 135–141, Jan. 2003.
- [52] S. Fidler, D. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 337–350, Mar. 2006.
- [53] A. Hanssen, A.-B. Salbergy, and L. L. Scharf, "Robust multidimensional matched subspace classifiers based on weighted least-squares," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 873–880, Mar. 2007.
- [54] K. C. Ho, P. D. Gader, J. N. Wilson, and T. Glenn, "Landmine detection using frequency domain features from GPR measurements and their fusion with time domain features," in *Proc. SPIE—Conf. Detection Remediation Technologies Mines Minelike Targets X*, Orlando, FL, Mar. 2005, pp. 1141–1150.



K. C. Ho (S'89–M'91–SM'00) received the B.Sc. degree with first-class honors in 1988 and the Ph.D. degree in 1991 from the Chinese University of Hongkong.

He was a Research Associate in the Royal Military College of Canada from 1991 to 1994. He joined Bell-Northern Research, Montreal, Canada, in 1995 as a member of the scientific staff. He was a Research Associate Professor in the Department of Electrical Engineering at the University of Saskatchewan, Saskatoon, Canada, from September 1996 to August 1997. Since September 1997, he has been with the University of Missouri-Columbia, Columbia, where he is currently a Professor in the Electrical and Computer Engineering Department. His research interests are in source localization, statistical signal processing, subsurface object detection, wavelet transform, wireless communications, and the development of efficient adaptive signal processing algorithms for various applications including landmine detection, echo cancellation, and time delay estimation. He has been active in the

development of the *ITU Standard Recommendation G.168* since 1995. He is the editor of the *ITU Standard Recommendations G.168: Digital Network Echo Cancellers* and *G.160: Voice Enhancement Devices for Mobile Networks*. He holds or coholds 12 patents.

Dr. Ho received the Junior Faculty Research Award from College of Engineering of the University of Missouri-Columbia in 2003. He has served as an Associate Editor of the *IEEE TRANSACTIONS ON SIGNAL PROCESSING* from 2003 to 2006, and the *IEEE SIGNAL PROCESSING LETTERS* from 2004 to 2008.

Lawrence Carin (SM'96–F'01) was born March 25, 1963, in Washington, DC. He received the B.S., M.S., and Ph.D. degrees in electrical engineering at the University of Maryland, College Park, in 1985, 1986, and 1989, respectively.

In 1989 he joined the Electrical Engineering Department at Polytechnic University, New York, as an Assistant Professor, and became an Associate Professor there in 1994. In September 1995, he joined the Electrical Engineering Department at Duke University, Durham, NC, where he is now the William H. Younger Professor of Engineering. He was the Principal Investigator (PI) on a Multidisciplinary University Research Initiative (MURI) on demining (1996–2001), and he is currently the PI of a MURI dedicated to multimodal inversion. He cofounded the Signal Innovations Group, Inc. (SIG), which is now a subsidiary of Integrian, Inc.; he serves as the Director of Technology at SIG. His current research interests include signal processing, sensing, and machine learning.

Dr. Carin is a member of the Tau Beta Pi and Eta Kappa Nu honor societies. He was an Associate Editor of the *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION* from 1995 to 2004.



Paul D. Gader (M'87–SM'99) received the Ph.D. degree in mathematics in 1986 from the University of Florida, Gainesville.

He was a Senior Research Scientist at Honeywell's Systems and Research Center, a Research Engineer and Manager at the Environmental Research Institute of Michigan, and a faculty member at the University of Wisconsin-Oshkosh and the University of Missouri-Columbia. He is currently a Professor in the Computer and Information Science and Engineering Department at the University of Florida. His research interests include signal and image analysis, pattern recognition, and machine learning. He has led teams involved in real-time, handwritten address recognition systems for the U.S. Postal Service. He has led teams that devised and tested several real-time algorithms in the field for mine detection and is currently involved in a variety of landmine detection projects involving algorithm development for single- and multisensor systems using GPRs, EM induction sensors, EO/IR imaging systems, and hyperspectral imaging sensors.



Joseph N. Wilson (M'05) received the B.S. degree in applied mathematics with the emphasis on computer science from the Florida State University, Tallahassee, in 1977 and the M.S. degree in applied mathematics and computer science and the Ph.D. degree in computer science from the University of Virginia, Charlottesville, in 1980 and 1985, respectively.

Since 1984, he has been a member of the faculty of the Computer and Information Science and Engineering Department at the University of Florida, Gainesville, where he served as Associate Chair from 1994 to 2001. His research interests include machine intelligence, image and signal processing, programming languages, and file systems.

Optimizing the Area Under a Receiver Operating Characteristic Curve With Application to Landmine Detection

Wen-Hsiung Lee, Paul D. Gader, *Senior Member, IEEE*, and Joseph N. Wilson

Abstract—A common approach to training neural network classifiers in a supervised learning setting is to minimize the mean-square error (mse) between the network output for each labeled training sample and some desired output. In the context of landmine detection and discrimination, although the performance of an algorithm is correlated with the mse, it is ultimately evaluated by using receiver operating characteristic (ROC) curves. In general, the larger the area under the ROC curve (AUC), the better. We present a new method for maximizing the AUC. Desirable properties of the proposed algorithm are derived and discussed that differentiate it from previously proposed algorithms. A hypothesis test is used to compare the proposed algorithm to an existing algorithm. The false alarm rate achieved by the proposed algorithm is found to be less than that of the existing algorithm with 95% confidence.

Index Terms—Area under the ROC curve (AUC), ground-penetrating radar (GPR), landmine detection, pattern recognition.

I. INTRODUCTION

LANDMINE detection algorithms often consist of two steps: a prescreener followed by an algorithm that discriminates between landmines and false alarms produced by the prescreener. We consider the discrimination portion of the landmine detection problem here.

Landmine detection algorithms are generally evaluated in terms of receiver operating characteristic (ROC) curves, which are parametric curves plotting the probability of detection (PD) against false alarm rate (FAR). Although the FAR is often given in terms of probability of false alarm, in landmine detection, it is often given in terms of the number of false alarms per square meter.

Several trainable algorithms have been applied to the problem of discrimination between landmines and false alarms, including hidden Markov models [1]–[3], neural networks [4]–[7], support and relevance vector machines [8], [9], fuzzy systems [10]–[12], and Choquet integrals [13], [14]. Algorithm parameters are usually estimated by optimizing objective func-

tions on training sets. Common objective functions include likelihood functions, mean-square error (mse), margin between classes, and minimum classification error.

Since the performance of a landmine detection algorithm is evaluated using ROC curves, it is logical to optimize the ROC curve in some sense. Objective functions for maximizing the area under the ROC curve (AUC) have been proposed in [15] and [16]. In this paper, we derive a new algorithm for training a differentiable two-class classifier that maximizes an objective function that approximates the AUC. We refer to the objective function as the ROCA objective function and the algorithm for training with respect to this objective function as the ROCA training algorithm. The ROCA objective function, unlike previously proposed objective functions, can be made arbitrarily close to the exact AUC. In addition, the algorithm has a significantly different behavior for false alarms with high confidence values, which we believe as advantageous for landmine detection. After deriving and analyzing ROCA, an application to landmine detection is presented. A specialized artificial neural network used for this problem previously [17], called the feedforward ordered weighted average (FOWA) network, was trained using ROCA, mse objective function, and the algorithm proposed by Yan *et al.* [15]. We will refer to the algorithm by Yan *et al.* as the WMW algorithm since its objective function is based on the Wilcoxon-Mann-Whitney statistic. To reduce the effects of random initialization, networks were trained 50 times, and the FARs for fixed PDs were averaged over the 50 training runs. Comparisons are made among ROCA, mse, and WMW algorithms. ROCA training outperformed the mse, producing average reductions in FAR between 44% and 56% for PD between 90% and 100%. By conducting hypothesis tests, we conclude that the ROCA algorithm also outperformed the WMW algorithm, producing statistically significant average reductions in FAR between 5% and 16% for PD between 90% and 100%.

II. ROC AREA OPTIMIZATION ALGORITHM

In this section, we derive the ROCA training algorithm and compare it analytically to the WMW algorithm. The ROCA algorithm is based on differentiation of a function related to the AUC. Yan *et al.* [15] and Rakotomamonjy [16] have noted that the exact AUC is nondifferentiable with respect to the classifier's parameters. However, while it is true that there is no real-valued function that is the derivative of the AUC, the derivative does exist in the sense of generalized functions. The

Manuscript received March 14, 2006; revised July 10, 2006. This work was supported in part by the Army Research Office under Grant W911NF-05-1-0067.

W.-H. Lee is with NIITEK Inc., Sterling, VA 20166 USA (e-mail: wlee@niitek.com).

P. D. Gader and J. N. Wilson are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: pgader@cise.ufl.edu; jnw@cise.ufl.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2006.887018

generalized function form of the derivative does not directly lead to a useful algorithm for optimizing the AUC but does lend some insight. A finite difference technique can be used to approximate the derivative. The approximate form converges to the generalized function form as the finite difference goes to zero. This approximation leads to a useful algorithm for optimizing the AUC. We show in this paper that, as long as the functional form of the classifier is differentiable with respect to its parameters, the approximate AUC can be optimized by gradient descent, and that the approximate AUC can be made arbitrarily close to the exact AUC. In Section II-A, we derive the ROCA algorithm. In Section II-B, we show analytically how the ROCA algorithm differs from the WMW algorithm. Experimental comparisons are given in Section III.

A. Algorithm Derivation

In this section, we first provide an expression for the AUC. We then derive an expression for the derivative of the AUC in terms of delta functions and point out why the expression does not yield a useful algorithm. Following that, we use finite differences to derive an expression for approximate AUC that can be optimized. The approximate AUC converges to the exact AUC as the finite difference goes to zero. We point out how the derivative of the approximate AUC parallels that of the exact AUC. Finally, we provide an algorithm for optimizing the approximate AUC using gradient descent.

We assume a set of training samples, $T = X \cup Y$, where $X = \{\mathbf{x}^i | i = 1, 2, \dots, M\}$ is a set of feature vectors computed from class 1 (e.g., landmines in our case) and $Y = \{\mathbf{y}^j | j = 1, 2, \dots, N\}$ is a set of feature vectors computed from class 2 (e.g., prescreener false alarms). We seek to use the training data to estimate the parameter vector θ for a system $f(\cdot; \theta)$ that maps an input feature vector \mathbf{z} from a training sample to a confidence $f(\mathbf{z}; \theta) \in [t_{\min}, t_{\max}]$ that \mathbf{z} represents a sample from class 1. The larger the value of $f(\mathbf{z}; \theta)$, the more likely that \mathbf{z} represents a sample from class 1. To simplify notation, define

$$d_{ij}(\theta) \equiv f(\mathbf{x}^i; \theta) - f(\mathbf{y}^j; \theta). \quad (1)$$

The PD and FAR at threshold value t , denoted by $P(t)$ and $F(t)$, respectively, are given by

$$P(t) = \frac{1}{M} \sum_{i=1}^M u(f(\mathbf{x}^i; \theta) - t) \quad (2)$$

$$F(t) = \frac{1}{A} \sum_{j=1}^N u(f(\mathbf{y}^j; \theta) - t) \quad (3)$$

where $u(a)$ is one if $a > 0$ and zero otherwise, and A represents either the total number of prescreener false alarms ($A = N$) or, as is often the case in mine detection, A is the total area of the lanes from which the training data are drawn. For the purpose of general development, A is just a constant. The exact AUC is

$$\text{AUC} = \int_0^\infty \text{PD}|_{\text{FAR}=F(t)} dF(t) \quad (4)$$

where $\text{PD}|_{\text{FAR}=F(t)}$ is the PD at $\text{FAR} = F(t)$ and $\text{PD}|_{\text{FAR}=F(t)} = P(t)$. The upper limit of the integral is set to ∞ to take into consideration all possibilities for A and N , including the extreme case where A is the total area and N is infinity. (Note that we will refer to the exact AUC as either exact AUC or AUC.)

Replacing $dF(t)$ with $F'(t)dt$ and using the expressions for $P(t)$ and $F(t)$ in (2) and (3) yield

$$\begin{aligned} \text{AUC} &= \int_{t_{\max}}^{t_{\min}} P(t) \frac{dF(t)}{dt} dt \\ &= \int_{t_{\max}}^{t_{\min}} \frac{1}{M} \sum_{i=1}^M u(f(\mathbf{x}^i; \theta) - t) \frac{d}{dt} \frac{1}{A} \sum_{j=1}^N u(f(\mathbf{y}^j; \theta) - t) dt \\ &= \frac{1}{MA} \sum_{i=1}^M \sum_{j=1}^N \int_{t_{\min}}^{t_{\max}} u(f(\mathbf{x}^i; \theta) - t) \delta(f(\mathbf{y}^j; \theta) - t) dt \\ &= \frac{1}{MA} \sum_{i=1}^M \sum_{j=1}^N u(d_{ij}(\theta)) \end{aligned} \quad (5)$$

where δ represents the Dirac delta function. Optimization using derivatives requires the derivative of the terms of the sums in (5). This derivative can be written in terms of delta functions as

$$\begin{aligned} \frac{\partial}{\partial \theta} \left[\frac{1}{MA} \sum_{i=1}^M \sum_{j=1}^N u(d_{ij}(\theta)) \right] \\ = \frac{1}{MA} \sum_{i=1}^M \sum_{j=1}^N \delta(d_{ij}(\theta)) \frac{\partial}{\partial \theta} d_{ij}(\theta). \end{aligned} \quad (6)$$

If we were to use (6) in a gradient-descent-based update formula, then parameter updates would only occur if $d_{ij}(\theta) = 0$. Since this is extremely unlikely to happen, this derivative does not lead directly to a useful algorithm. The problem can be remedied using a finite difference approximation of the derivative of F . This leads to a gradient-descent algorithm that updates parameters whenever $0 < d_{ij}(\theta) \leq \Delta t$, where Δt is the finite difference. To this end, let Δt denote a nonnegative real number. Then

$$\begin{aligned} \text{AUC} &= \int_{t_{\max}}^{t_{\min}} P(t) F'(t) dt \\ &\approx \int_{t_{\max}}^{t_{\min}} P(t) \frac{F(t) - F(t - \Delta t)}{\Delta t} dt \\ &= \int_{t_{\min}}^{t_{\max}} P(t) \frac{F(t - \Delta t) - F(t)}{\Delta t} dt. \end{aligned} \quad (7)$$

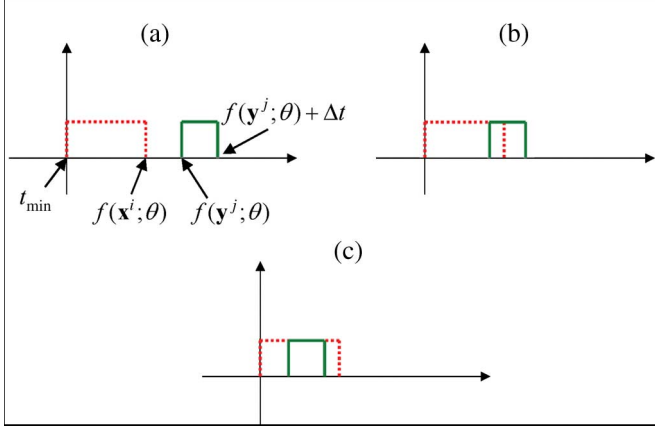


Fig. 1. Three overlap cases between the intervals $[t_{\min}, f(\mathbf{x}^i; \theta)]$ and $[f(\mathbf{y}^j; \theta), f(\mathbf{y}^j; \theta) + \Delta t]$.

Substituting the expressions for PD and FAR from (2) and (3) into (7) yields the following expression for the approximate AUC, which we refer to as $J(\Delta t)$:

$$\begin{aligned}
 J(\Delta t) &= \frac{1}{\Delta t} \int_{t_{\min}}^{t_{\max}} \left[\frac{1}{M} \sum_{i=1}^M u(f(\mathbf{x}^i; \theta) - t) \right] \\
 &\quad \cdot \left[\frac{1}{A} \sum_{j=1}^N u(f(\mathbf{y}^j; \theta) - (t - \Delta t)) \right. \\
 &\quad \left. - \frac{1}{A} \sum_{j=1}^N u(f(\mathbf{y}^j; \theta) - t) \right] dt \\
 &= \frac{1}{\Delta t \cdot MA} \sum_{i=1}^M \sum_{j=1}^N \int_{t_{\min}}^{t_{\max}} u(f(\mathbf{x}^i; \theta) - t) \\
 &\quad \cdot [u(f(\mathbf{y}^j; \theta) - (t - \Delta t)) - u(f(\mathbf{y}^j; \theta) - t)] dt. \quad (8)
 \end{aligned}$$

Clearly, $\lim_{\Delta t \rightarrow 0} J(\Delta t) = \text{AUC}$. We also show in Appendix A that $J(\Delta t)$ approaches AUC monotonically from below.

Note that the terms $u(f(\mathbf{x}^i; \theta) - t)$ and $[u(f(\mathbf{y}^j; \theta) - (t - \Delta t)) - u(f(\mathbf{y}^j; \theta) - t)]$ of the integrand in (8) are nonzero only if $t_{\min} \leq t \leq f(\mathbf{x}^i; \theta)$ and $f(\mathbf{y}^j; \theta) \leq t \leq f(\mathbf{y}^j; \theta) + \Delta t$, respectively. For the integrand to be nonzero for a pair i and j , the two intervals $[t_{\min}, f(\mathbf{x}^i; \theta)]$ and $[f(\mathbf{y}^j; \theta), f(\mathbf{y}^j; \theta) + \Delta t]$ must overlap. The calculation of the integral in (8) can be split into three cases corresponding to different types of overlap as shown in Fig. 1.

To write these cases out, let

$$\begin{aligned}
 C_1 &= \{(i, j) | d_{ij}(\theta) \leq 0\} \\
 C_2 &= \{(i, j) | d_{ij}(\theta) \in (0, \Delta t)\} \\
 C_3 &= \{(i, j) | d_{ij}(\theta) > \Delta t\}. \quad (9)
 \end{aligned}$$

Also, denote the terms in the sum in (8) as T_{ij} . Then

$$\begin{aligned}
 T_{ij} &\equiv \int_{t_{\min}}^{t_{\max}} u(f(\mathbf{x}^i; \theta) - t) \\
 &\quad \times [u(f(\mathbf{y}^j; \theta) - (t - \Delta t)) - u(f(\mathbf{y}^j; \theta) - t)] dt \\
 &= \begin{cases} \text{(a)} \ 0, & \text{if } (i, j) \in C_1 \\ \text{(b)} \ \int_{f(\mathbf{y}^j; \theta)}^{f(\mathbf{x}^i; \theta)} 1 \cdot dt = d_{ij}(\theta), & \text{if } (i, j) \in C_2 \\ \text{(c)} \ \int_{f(\mathbf{y}^j; \theta)}^{f(\mathbf{y}^j; \theta) + \Delta t} 1 \cdot dt = \Delta t, & \text{if } (i, j) \in C_3. \end{cases} \quad (10)
 \end{aligned}$$

The cases in (10) are depicted in Fig. 1. The objective function $J(\Delta t)$ is expressed as a function of $f(\mathbf{x}^i; \theta)$, $i = 1, 2, \dots, M$, and $f(\mathbf{y}^j; \theta)$, $j = 1, 2, \dots, N$, so the gradient descent can be used to iteratively update the parameter vector θ as long as the derivative of the function $f(\cdot; \theta)$ with respect to θ can be determined. To compute the update term for θ , the update terms due to every pair of \mathbf{x}^i and \mathbf{y}^j are accumulated until all pairs of \mathbf{x}^i and \mathbf{y}^j are processed and θ is updated. This process is repeated until some convergence criterion is met. More precisely, the ROCA training algorithm proceeds as follows:

Algorithm: ROCA training

Initialize θ

Do until stopping criterion reached

1. Compute $f(\mathbf{x}^i; \theta)$ and $f(\mathbf{y}^j; \theta)$ for every training sample

2. Set $\nabla_{\theta}^{ave} J(\Delta t) = 0$

3. For each pair of training samples \mathbf{x}^i and \mathbf{y}^j

- Identify which case, (a)–(c), of (10) is satisfied

- If case (b) is satisfied then

Compute $\nabla_{\theta}^{i,j} J(\Delta t) = (\partial J(\Delta t) / \partial f)(\partial f / \partial \theta)$ using (b)

Set $\nabla_{\theta}^{ave} J(\Delta t) = \nabla_{\theta}^{ave} J(\Delta t) + \nabla_{\theta}^{i,j} J(\Delta t)$

End If

End For

4. Update $\theta = \theta + \eta \nabla_{\theta}^{ave} J(\Delta t)$

End Do

Given the training set $T = X \cup Y$ and a classifier f , let D be defined as $D \equiv \min_{i,j} \{\max(d_{ij}(\theta), 0)\}$. Then, if $0 < \Delta t < D$, the objective function $J(\Delta t)$ in (8) always yields the AUC in (5). The only exception is when $D = 0$ and AUC is 0. Two other desirable properties of $J(\Delta t)$ are presented in Appendix A. Specifically, we show that $J(\Delta t)$ is monotonic and $\lim_{\Delta t \rightarrow 0} J(\Delta t) = \text{AUC}$. The limit is reached for $0 < \Delta t < D$ as indicated above.

The method for deciding on a value for Δt offers opportunities for future research. Clearly, different choices lead to different behaviors. For example, if $0 < \Delta t < D$, all derivatives will be zero and no update will occur. If Δt is too large, then we use all pairs for updating, even those for which

$f(\mathbf{x}^i; \theta) \gg f(\mathbf{y}^j; \theta)$. Furthermore, one could choose a single value for Δt , or devise a method by which Δt is modified every training iteration. There are many possibilities. In our experiments, we heuristically chose $\Delta t = (\min_j (f(\mathbf{y}^j; \theta)) - t_{\min})/2$ as we found that such a Δt was not too small nor too large. Future work will focus on more analytical method for choosing Δt .

B. Analytical Comparison to WMW Algorithm

By contrast, consider the WMW algorithm by Yan *et al.* [15]. The WMW algorithm minimizes the following objective function:

$$U_R(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta)) = \sum_{i=1}^M \sum_{j=1}^N R_1(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta)) \quad (11)$$

where

$$R_1(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta)) = \begin{cases} -(d_{ij}(\theta) - \gamma)^p, & \text{if } d_{ij}(\theta) < \gamma \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

with $0 < \gamma \leq 1$ and $p > 1$.

There are several differences between the objective function U_R and $J(\Delta t)$. To analyze the differences, first notice that U_R is formulated as a minimization rather than a maximization problem. We can change the WMW objective function into a maximization problem by posing it as a maximization of $B - U_R/MA$, where B is the maximum possible value of AUC. Now, suppose the confidence value of each mine is higher than the confidence value of every false alarm. Then, the AUC reaches its maximum value at $B = (M \cdot N)/(M \cdot A) = N/A$. The objective function $U_R(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta))$ is equal to $(B - \text{AUC})MA$ only when $p = 0$ and $\gamma = 0$. But, the objective function is a constant when $p = 0$ and cannot be optimized. In addition, although $\lim_{p \rightarrow 0, \gamma \rightarrow 0} U_R = (B - \text{AUC})MA$, the value $p = 0$ is also not allowed, since p must be larger than one according to the definition in [15]. On the other hand, if $p \neq 0$, it is not clear how $U_R(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta))$ is related to AUC, since $U_R(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta))$ depends explicitly on the difference of $f(\mathbf{x}^i; \theta)$ and $f(\mathbf{y}^j; \theta)$, $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$, but AUC depends only on whether $f(\mathbf{x}^i; \theta)$ is greater than $f(\mathbf{y}^j; \theta)$ for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$.

To see the impact of this property, suppose there are outliers in the training data such that $d_{ij}(\theta) \ll 0$ for some pair i and j (the confidence for the i th mine is much less than that of the j th false alarm). This can happen if the i th mine has a very poor signature and the j th false alarm has a very strong signature. It is unwise to train against such false alarms, but when training against large databases, it is difficult to screen out such examples. While the ROCA algorithm does not update the parameters when encountering such pairs [see case (a) in (10)], those pairs dominate the parameter updates in the WMW algorithm because the first factor in the expressions for the

derivatives of the terms of the objective function will be large, resulting in large changes in parameters. The derivative is

$$\frac{\partial R_1(f(\mathbf{x}^i; \theta), f(\mathbf{y}^j; \theta))}{\partial \theta} = \begin{cases} p(-(d_{ij}(\theta) - \gamma))^{p-1} \\ \times \left(-\frac{\partial f(\mathbf{x}^i; \theta)}{\partial \theta} + \frac{\partial f(\mathbf{y}^j; \theta)}{\partial \theta} \right), & \text{if } d_{ij}(\theta) < \gamma \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

That is, the WMW algorithm is forcing the parameters to recognize the outliers. In the problem of landmine detection, for which we will present experimental results in the next section, it is not unusual to have outliers like large metal objects, which have strong ground-penetrating radar (GPR) signals. To force the parameters to learn the characteristics of outliers is often undesirable. In this aspect, the ROCA algorithm is more immune to outliers and therefore potentially more robust than the WMW algorithm.

Picking from values suggested by Yan *et al.* [15], we tried various values for p and γ in experiments for the problem of landmine detection. The results for $p = 2$ and $\gamma = 0.1$ are presented in the next section as those values lead to the best performance of the WMW algorithm.

III. APPLICATION TO LANDMINE DETECTION

We provide an example of optimizing the AUC for the problem of discriminating between landmines and false alarms using a classifier called FOWA. The FOWA network has been described in previous publications [17] and satisfies the requirements of the system $f(\cdot; \theta)$ described in the previous section. In the context of landmine detection, given a feature vector \mathbf{z} , the FOWA network computes a single value $f(\mathbf{z}; \theta)$ that can be interpreted as the confidence that \mathbf{z} represents a mine. The FOWA network is a standard feedforward (F) network coupled with a unique front end that integrates values over depth using ordered weighted averaging (OWA) operators. The elements of θ are the weights of the feedforward network and the OWA operators. For completeness, the training algorithm is given in Appendix B.

The mse, ROCA, and WMW algorithms were evaluated using the GPR data collected from outdoor test lanes at two different locations. The first, Site 1, is in a temperate region with significant rainfall, whereas the second, Site 2, is in an arid region. Soil was moist at Site 1 and very dry at Site 2. The lanes at both sites are simulated dirt or gravel roads. The lanes at Site 1 are 500 m long and 3 m wide. The lanes at Site 2 are 300 m long and 3 m wide. The lanes at Site 1 are labeled 1A, 1B, and 1C, and contain mines. The lanes at Site 2 are labeled lanes 2A, 2B, and 2C. Lanes 2A and 2B contain mines, and lane 2C contains both mines and emplaced clutter items, such as pieces of metal and wood. The numbers of mines of each type in the lanes are given in Table I. All the mines are Anti-Tank mines.

Two data collections were performed at each site resulting in a total of four collections, each in a different month. The collections at Site 1 were made in November and December 2002, and the collections at Site 2 were made in October

TABLE I
NUMBERS OF MINES, CLUTTERS, AND HOLES IN CALIBRATION LANES

	Lane 1A	Lane 1B	Lane 1C	Lane 2A	Lane 2B	Lane 2C	All lanes
Metal	4	4	4	4	4	0	20
Plastic	14	14	16	10	10	11	75
Clutter	0	0	0	0	0	32	32
Hole	0	0	0	0	0	7	7
Total mines	18	18	20	14	14	11	95

2002 and January 2003. A prescreener [18] was run on all the lanes creating a set of 170 mines detected and 978 prescreener false alarms. Altogether, 172 mine encounters are possible. The prescreener missed two mines, and therefore, the highest PD is $170/172 = 0.9884$. With the total area of the mine lanes at 7213.98 m^2 , one false alarm contributes to $1.39 \times 10^{-4} (/ \text{m}^2)$ of FAR. This set was used for comparing the mse, ROCA, and WMW algorithms. Geometric features including eccentricity, solidity, compactness, ratio of area to filled area [19] are computed for each of the alarms. Those geometric features along with the prescreener output value are used as input to the FOWA network.

To compare the mse, ROCA, and WMW algorithms for training the FOWA network, their performance must be evaluated statistically because randomly initialized weights are involved. As is the case with almost all training algorithms, different initializations may lead to different solutions because of local extrema. Hence, the average performance must be calculated over a set of training experiments, each using a different initialization. Therefore, 50 training experiments or runs were conducted for the mse, ROCA, and WMW algorithms. Each run consisted of lane-based cross validation for each of the mse, ROCA, and WMW algorithms. Lane-based cross validation is described by the following pseudocode.

```

Algorithm: Lane-based cross validation
For each lane L
    Validation set = {set of all alarms from all
                     occurrences of lane L}
    Training set = {set of all alarms from all
                   occurrences of other lanes}
    Train on training set until stopping criterion met
    Assign confidence values to all alarms in validation set
End For
Generate ROC curve based on confidence values of all
alarms from all lanes.

```

For the mse approach, the FOWA weights were trained by minimizing the mse of the training data with -0.7 and 0.7 as the desired values of the classes of mines and nonmines, respectively. These desired values were previously found to perform well [17]. The AUC for the cross-validation data was used as the criterion for picking the set of trained weights to make sure that the final weights led to a maximum AUC for the cross-validation data. Using the cross-validation data for determining when to stop the training prevents overfitting. By using the weights thus obtained, we recorded the scores for the cross-validation data. Since the ultimate goal is to improve upon the FAR obtained by the mse algorithm in each run, the weights obtained by the mse algorithm were then used as the initial weights for both the ROCA and WMW algorithms. For

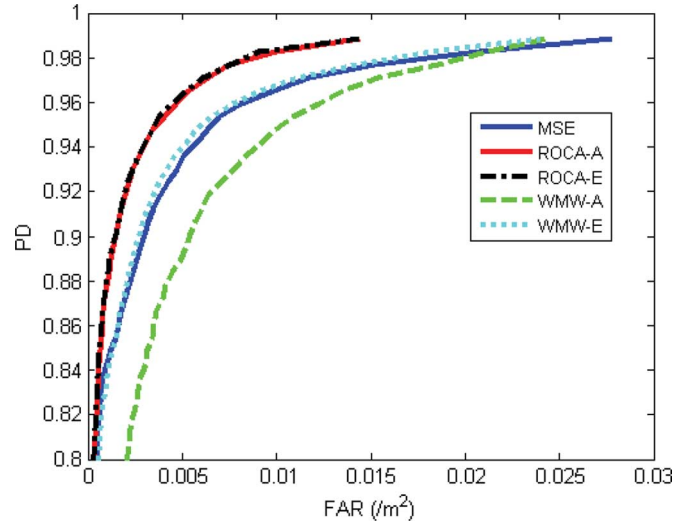


Fig. 2. Average ROC curves over 50 runs.

the ROCA algorithm, two approaches were taken to determine when to stop the training: the first one (referred to as ROCA-A, i.e., ROCA with Approximate AUC) used the approximate AUC in (8), and the second one (referred to as ROCA-E, i.e., ROCA with Exact AUC) used the exact AUC in (5). For the WMW algorithm, two approaches were also taken to determine when to stop the training: The first one (referred to as WMW-A) used U_R , and the second one (referred to as WMW-E) used the exact AUC in (5). For every algorithm, the FOWA network has five OWA operators, one input layer with eight nodes, one output layer with single output, and one hidden layer with 15 hidden nodes.

For a given threshold, a mine is considered detected if there is an alarm within 0.25 m from the edge of the mine with confidence value above the threshold. Given a threshold, the PD for a lane or set of lanes is defined to be the number of mines detected divided by the number of mines. A false alarm is an alarm with confidence above the threshold and with location farther than 0.25 m from the edge of any mine. The FAR is defined as the number of false alarms per square meter.

Fig. 2 shows the average ROC curves over 50 runs for mse, ROCA-A, ROCA-E, WMW-A, and WMW-E. Since the approximate AUC in (8) can be arbitrarily close to the exact AUC in (5), the difference between the ROC curves of ROCA-A and ROCA-E is minimal as expected. On the other hand, WMW-A does not perform nearly as well as WMW-E. WMW-E shows a slight improvement on FAR over the mse algorithm.

Table II shows the average score (PD versus FAR) over 50 runs for the mse, ROCA-E, and WMW-E algorithms. Also shown in Table II is the percentage reduction of FAR by the ROCA-E and WMW-E algorithms over the mse algorithm. For PD higher than 90% and attainable (100% PD was not attained), the reduction achieved by the ROCA-E algorithm is at least 44%. On the other hand, the largest reduction achieved by the WMW-E at any one level of PD is 15.8%.

To determine whether the difference between the average scores of ROCA-E and WMW-E is significant, we ran hypothesis tests on the FAR at each value of PD above 90%. Assuming

TABLE II
AVERAGE FAR OVER 50 EXPERIMENTS

PD	FAR(MSE)	WMW-E		ROCA-E	
		FAR	Reduction (%) over FAR(MSE)	FAR	Reduction (%) over FAR(MSE)
0.8953	0.0028	0.0025	10.2	0.0013	53.8
0.9012	0.0030	0.0027	10.4	0.0015	50.2
0.9070	0.0032	0.0029	8.9	0.0016	48.9
0.9128	0.0035	0.0032	9.8	0.0018	48.9
0.9186	0.0038	0.0035	9.2	0.0019	50.3
0.9244	0.0042	0.0038	12.7	0.0021	50.2
0.9302	0.0047	0.0042	13.1	0.0024	48.9
0.9360	0.0051	0.0046	10.6	0.0027	48.0
0.9419	0.0057	0.0052	10.7	0.0030	47.6
0.9477	0.0063	0.0057	10.6	0.0034	46.2
0.9535	0.0070	0.0065	6.4	0.0038	45.6
0.9593	0.0081	0.0077	5.3	0.0045	44.9
0.9651	0.0098	0.0091	7.6	0.0052	46.7
0.9709	0.0117	0.0112	5.1	0.0062	47.6
0.9767	0.0152	0.0143	6.2	0.0076	49.9
0.9826	0.0208	0.0183	13.3	0.0091	56.1
0.9884	0.0278	0.0240	15.8	0.0144	48.3

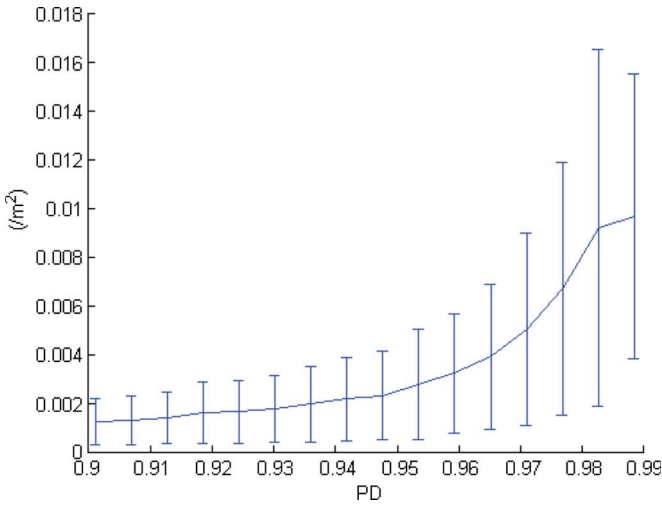


Fig. 3. Plot of 95% confidence interval for (mean FAR of WMW-E - mean FAR of ROCA-E) versus PD.

the FAR at each value of PD is normally distributed, it was shown that by conducting two-sample T-tests that with 95% confidence, the mean FAR of WMW-E is always higher than that of ROCA-E. Fig. 3 shows the 95% confidence interval for the difference between mean FAR of WMW-E and FAR of ROCA-E at each value of PD higher than 90%.

IV. CONCLUSION

The ROCA algorithm has been derived for maximizing the approximate AUC. The ROCA algorithm is general; therefore, it can be used to train weights for any systems with functional forms that are differentiable with respect to the parameters. A specific application to landmine detection using FOWA networks was described. Experiments show that the ROCA algorithm reduces the FAR over the mse algorithm by 44%–56% for PDs in the range of 90%–100%. By contrast, the previously proposed WMW algorithm reduces the FAR over the mse algorithm by 5.1%–15.8% for the same range of PD. The ROCA algorithm outperformed the WMW algorithm, and the difference is statistically significant with 95%

confidence. An analysis of the update equations shows that the WMW algorithm treats outliers quite differently than the ROCA algorithm, and this difference is likely to produce the improved performance demonstrated by the ROCA algorithm.

One may wish to restrict the range of PD or FAR and optimize the area under a portion of the ROC curve. This restriction would require restricting the values of the threshold variable t dynamically, since the specific values that yield the correct interval over which to optimize would change from iteration to iteration. For example, if one wishes to optimize only for values of FAR in the interval $[F_{\text{low}}, F_{\text{high}}]$, then at each iteration values of t , say t_{low} and t_{high} , must be found so that $F(t_{\text{low}}) = F_{\text{low}}$ and $F(t_{\text{high}}) = F_{\text{high}}$. Restrictions on the values of t lead to several additional cases that need to be added to (10) depending how the values $f(\mathbf{x}^i, \theta)$ and $f(\mathbf{y}^j, \theta)$ compare to the values t_{low} and t_{high} . While it is straightforward to write out the cases, the increased level of detail would unnecessarily obscure the current discussion. It is also possible to express P and F as functions of some appropriate $h(t, \theta)$ rather than simply t , as we have chosen to do. Such a change could support more complex and perhaps robust criteria for specifying those aspects of the ROC to be optimized. It is an interesting subject for further research to investigate whether or not such changes would lead to an enhanced performance.

APPENDIX A

PROPERTIES OF THE PROPOSED OBJECTIVE FUNCTION

Property 1. $J(\Delta t) \leq \text{AUC}$:

Proof:

$$\begin{aligned}
 J(\Delta t) &= \frac{1}{\Delta t \cdot MA} \sum_{i=1}^M \sum_{j=1}^N T_{ij} \\
 &= \frac{1}{\Delta t \cdot MA} \sum_{(i,j) \in C_1} 0 + \frac{1}{\Delta t \cdot MA} \sum_{(i,j) \in C_2} d_{ij}(\theta) \\
 &\quad + \frac{1}{\Delta t \cdot MA} \sum_{(i,j) \in C_3} \Delta t \\
 &\leq \sum_{(i,j) \in C_2} \frac{\Delta t}{\Delta t \cdot MA} + \sum_{(i,j) \in C_3} \frac{1}{MA} = \frac{1}{MA} |C_2 \cup C_3|.
 \end{aligned}$$

Note that $(i, j) \in C_2 \cup C_3$ if and only if $u(d_{ij}(\theta)) = 1$. Thus, $J(\Delta t) \leq 1/MA \cdot \sum_{i,j} u(d_{ij}(\theta)) = \text{AUC}$. ■

Property 2: If $\Delta t_1 < \Delta t_2$, then $|\text{AUC} - J(\Delta t_1)| \leq |\text{AUC} - J(\Delta t_2)|$.

Proof: Note that by Property 1, $|\text{AUC} - J(\Delta t_k)| = \text{AUC} - J(\Delta t_k)$ for $k = 1, 2$. Furthermore

$$\begin{aligned}
 \text{AUC} - J(\Delta t_k) &= \sum_{(i,j) \in C_2^{(k)}} \frac{1}{MA} - \sum_{(i,j) \in C_2^{(k)}} \frac{d_{ij}(\theta)}{\Delta t_k \cdot MA} \\
 &= \frac{1}{MA} \sum_{(i,j) \in C_2^{(k)}} \frac{\Delta t_k - d_{ij}(\theta)}{\Delta t_k}
 \end{aligned}$$

where $C_2^{(1)}$ and $C_2^{(2)}$ correspond to the set C_2 for Δt_1 and Δt_2 , respectively. Note that $C_2^{(1)} \subseteq C_2^{(2)}$. Hence

$$\begin{aligned}
& |AUC - J(\Delta t_2)| - |AUC - J(\Delta t_1)| \\
&= (AUC - J(\Delta t_2)) - (AUC - J(\Delta t_1)) \\
&= \frac{1}{MA} \left[\sum_{(i,j) \in C_2^{(2)}} \frac{\Delta t_2 - d_{ij}(\theta)}{\Delta t_2} - \sum_{(i,j) \in C_2^{(1)}} \frac{\Delta t_1 - d_{ij}(\theta)}{\Delta t_1} \right] \\
&= \frac{1}{MA} \left[\sum_{(i,j) \in C_2^{(2)} - C_2^{(1)}} \frac{\Delta t_2 - d_{ij}(\theta)}{\Delta t_2} \right. \\
&\quad \left. + \sum_{(i,j) \in C_2^{(1)}} \left(\frac{\Delta t_2 - d_{ij}(\theta)}{\Delta t_2} - \frac{\Delta t_1 - d_{ij}(\theta)}{\Delta t_1} \right) \right] \\
&= \frac{1}{MA} \left[\sum_{(i,j) \in C_2^{(2)} - C_2^{(1)}} \frac{\Delta t_2 - d_{ij}(\theta)}{\Delta t_2} \right. \\
&\quad \left. + \sum_{(i,j) \in C_2^{(1)}} \frac{d_{ij}(\theta)(\Delta t_2 - \Delta t_1)}{\Delta t_1 \Delta t_2} \right] \geq 0.
\end{aligned}$$

APPENDIX B

DERIVATION OF FOWA ROCA TRAINING

As mentioned above, the system $f(\cdot; \theta)$ can assume any form as long as the derivative of $f(\cdot; \theta)$ with respect to θ can be determined. We have employed a FOWA neural network [17] as $f(\cdot; \theta)$. Extension of the proposed algorithm to other applicable $f(\cdot; \theta)$ is straightforward. Suppose the information we have for a training sample includes vectors of features $\alpha_m = [\alpha_{m,1}, \alpha_{m,2}, \dots, \alpha_{m,K_m}]^T$, $m = 1, 2, \dots, I_o$, with elements that we wish to sort first and scalar features $\alpha_{I_o+1}, \alpha_{I_o+2}, \dots, \alpha_I$. The input to the FOWA system is the whole collection of features $\mathbf{z} = [\alpha_1^T, \alpha_2^T, \dots, \alpha_{I_o}^T, \alpha_{I_o+1}, \alpha_{I_o+2}, \dots, \alpha_I]^T$ (see Fig. 4). Let the m th OWA [20]–[22] system have a weight vector $[w_{m,1}, w_{m,2}, \dots, w_{m,K_m}]^T$, and let $[\alpha_{m,1}, \alpha_{m,2}, \dots, \alpha_{m,K_m}]^T$ be the input vector of K_m features to the m th OWA system. Then, the output of the m th OWA system is

$$\lambda_m = \sum_{k=1}^{K_m} w_{m,k} \alpha_{m(k)}, \quad m = 1, 2, \dots, I_o \quad (14)$$

where I_o is the number of OWA systems and $\alpha_{m(k)}$ is the k th order statistic of the vector $[\alpha_{m,1}, \alpha_{m,2}, \dots, \alpha_{m,K_m}]^T$, i.e., $\alpha_{m(1)} \leq \alpha_{m(2)} \leq \dots \leq \alpha_{m(K_m)}$. For $m = I_o + 1, I_o + 2, \dots, I$, $\lambda_m = \alpha_m$, that is, these features are not sorted and

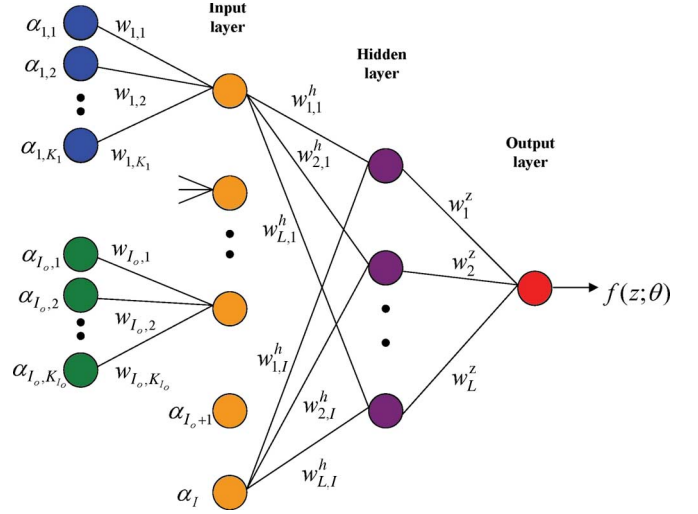


Fig. 4. FOWA network.

weighted by OWA operators. To satisfy the following constraints for the OWA systems

$$\sum_{k=1}^{K_m} w_{m,k} = 1, \quad m = 1, 2, \dots, I_o \quad (15)$$

$$0 \leq w_{m,k} \leq 1, \quad k = 1, 2, \dots, K_m, \quad m = 1, 2, \dots, I_o \quad (16)$$

the weights $\{w_{m,k}\}$ are implemented as

$$w_{m,k} = \frac{\nu_{m,k}^2}{\sum_{k=1}^{K_m} \nu_{m,k}^2}, \quad k = 1, 2, \dots, K_m, \quad m = 1, 2, \dots, I_o. \quad (17)$$

With tanh sigmoid functions being employed at the hidden and output layers, the outputs at the hidden layer and output layer are, respectively

$$h_l = \tanh \left(\beta_1 \sum_{m=1}^I w_{l,m}^h \lambda_m \right), \quad l = 1, 2, \dots, L \quad (18)$$

$$f(\mathbf{z}; \theta) = \tanh \left(\beta_2 \sum_{l=1}^L w_l^z h_l \right) \quad (19)$$

where L is the number of hidden nodes and θ is a vector with all the weights $\{w_{m,k}\}$, $\{w_{l,m}^h\}$, and $\{w_l^z\}$ as its elements. In our notation, \mathbf{z} can be either \mathbf{x}^i for mines or \mathbf{y}^j for nonmines.

Suppose for the training samples \mathbf{x}^i and \mathbf{y}^j , the values of λ_m are $\lambda_m^{(i)}$ and $\lambda_m^{(j)}$, respectively, and the values of h_l are $h_l^{(i)}$ and $h_l^{(j)}$, respectively. Instead of minimizing the mse between $f(\mathbf{x}^i; \theta)$ and its desired output and the mse between $f(\mathbf{y}^j; \theta)$ and its desired output as in [17], the proposed algorithm aims to maximize the objective function $J(\Delta t)$ in (8). To adaptively

update the weights θ , the steepest descent method is used. The incremental update for w_l^z , $l = 1, 2, \dots, L$, is

$$\begin{aligned} dw_l^z &= s_1 \frac{\partial J(\Delta t)}{\partial w_l^z} = \frac{s_1}{\Delta t \cdot MA} \sum_{i=1}^M \sum_{j=1}^N \frac{\partial T_{ij}}{\partial w_l^z} \\ &= \frac{s_1}{\Delta t \cdot MA} \sum_{(i,j) \in C_2} \left(\frac{\partial f(\mathbf{x}^i; \theta)}{\partial w_l^z} - \frac{\partial f(\mathbf{y}^j; \theta)}{\partial w_l^z} \right) \end{aligned} \quad (20)$$

where s_1 is a step size. The partial derivatives in (20) are equal to

$$\frac{\partial f(\mathbf{x}^i; \theta)}{\partial w_l^z} = (1 - f(\mathbf{x}^i; \theta)^2) \beta_2 h_l^{[i]} \quad (21)$$

$$\frac{\partial f(\mathbf{y}^j; \theta)}{\partial w_l^z} = (1 - f(\mathbf{y}^j; \theta)^2) \beta_2 h_l^{(j)}. \quad (22)$$

The incremental update for $w_{l,m}^h$, $m = 1, 2, \dots, I$, $l = 1, 2, \dots, L$, is

$$\begin{aligned} dw_{l,m}^h &= s_2 \frac{\partial J(\Delta t)}{\partial w_{l,m}^h} = \frac{s_2}{\Delta t \cdot MA} \sum_{i=1}^M \sum_{j=1}^N \frac{\partial T_{ij}}{\partial w_{l,m}^h} \\ &= \frac{s_2}{\Delta t \cdot MA} \sum_{(i,j) \in C_2} \left(\frac{\partial f(\mathbf{x}^i; \theta)}{\partial w_{l,m}^h} - \frac{\partial f(\mathbf{y}^j; \theta)}{\partial w_{l,m}^h} \right) \\ &= \frac{s_2}{\Delta t \cdot MA} \\ &\quad \times \sum_{(i,j) \in C_2} \left(\frac{\partial f(\mathbf{x}^i; \theta)}{\partial h_l^{[i]}} \frac{\partial h_l^{[i]}}{\partial w_{l,m}^h} - \frac{\partial f(\mathbf{y}^j; \theta)}{\partial h_l^{(j)}} \frac{\partial h_l^{(j)}}{\partial w_{l,m}^h} \right) \\ &= \frac{s_2}{\Delta t \cdot MA} \\ &\quad \times \sum_{(i,j) \in C_2} \left[(1 - f(\mathbf{x}^i; \theta)^2) \beta_2 w_l^z \left(1 - (h_l^{[i]})^2 \right) \right. \\ &\quad \times \beta_1 \lambda_m^{[i]} - (1 - f(\mathbf{y}^j; \theta)^2) \beta_2 w_l^z \\ &\quad \times \left. \left(1 - (h_l^{(j)})^2 \right) \beta_1 \lambda_m^{(j)} \right]. \end{aligned} \quad (23)$$

The incremental update for $\nu_{m,k}$, $k = 1, 2, \dots, K_m$, $m = 1, 2, \dots, I$, is

$$\begin{aligned} d\nu_{m,k} &= s_3 \frac{\partial J(\Delta t)}{\partial \nu_{m,k}} \\ &= \frac{s_3}{\Delta t \cdot MA} \sum_{i=1}^M \sum_{j=1}^N \frac{\partial T_{ij}}{\partial \nu_{m,k}} \end{aligned}$$

$$\begin{aligned} &= \frac{s_3}{\Delta t \cdot MA} \\ &\quad \times \sum_{(i,j) \in C_2} \left\{ \left(\sum_{l=1}^L \frac{\partial f(\mathbf{x}^i; \theta)}{\partial h_l^{[i]}} \frac{\partial h_l^{[i]}}{\partial \lambda_m^{[i]}} \right) \frac{\partial \lambda_m^{[i]}}{\partial w_{m,k}} \frac{\partial w_{m,k}}{\partial \nu_{m,k}} \right. \\ &\quad \left. - \left(\sum_{l=1}^L \frac{\partial f(\mathbf{y}^j; \theta)}{\partial h_l^{(j)}} \frac{\partial h_l^{(j)}}{\partial \lambda_m^{(j)}} \right) \frac{\partial \lambda_m^{(j)}}{\partial w_{m,k}} \frac{\partial w_{m,k}}{\partial \nu_{m,k}} \right\} \\ &= \frac{s_3}{\Delta t \cdot MA} \\ &\quad \times \left\{ \sum_{(i,j) \in C_2} \left\{ \left[\sum_{l=1}^L (1 - f(\mathbf{x}^i; \theta)^2) \right. \right. \right. \\ &\quad \times \beta_2 w_l^z \left(1 - (h_l^{[i]})^2 \right) \beta_1 w_{l,m}^h \left. \right] \\ &\quad \cdot x_{m,k}^i \left(\frac{2\nu_{m,k}}{\sum_{d=1}^{K_m} \nu_{m,d}^2} - \frac{\nu_{m,k}^2 \cdot 2\nu_{m,k}}{\left(\sum_{d=1}^{K_m} \nu_{m,d}^2 \right)^2} \right) \\ &\quad - \left[\sum_{l=1}^L (1 - f(\mathbf{y}^j; \theta)^2) \right. \\ &\quad \times \beta_2 w_l^z \left(1 - (h_l^{(j)})^2 \right) \beta_1 w_{l,m}^h \left. \right] \\ &\quad \cdot y_{m,k}^j \left(\frac{2\nu_{m,k}}{\sum_{d=1}^{K_m} \nu_{m,d}^2} \right. \\ &\quad \left. \left. - \frac{\nu_{m,k}^2 \cdot 2\nu_{m,k}}{\left(\sum_{d=1}^{K_m} \nu_{m,d}^2 \right)^2} \right) \right\} \right\}. \end{aligned} \quad (24)$$

The parameters s_2 and s_3 in (23) and (24) are step sizes, and the quantities $x_{m,k}^i$ and $y_{m,k}^j$ in (24) are the values of the feature $\alpha_{m,k}$ for the training samples \mathbf{x}^i and \mathbf{y}^j , respectively.

ACKNOWLEDGMENT

The authors would like to thank R. Harmon, R. Weaver, and P. Howard for their support of this work, and also to L. Carin, L. Collins, and P. Torrione (Duke University) and F. Clodfelter and others from NIITEK, Inc. for their technical discussions, insights, cooperation, and GPR data.

REFERENCES

- [1] P. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1231–1244, Jun. 2001.
- [2] Y. Zhao, P. Gader, P. Chen, and Y. Zhang, "Training DHMMs of mine and clutter to minimize landmine detection errors," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1016–1024, May 2003.
- [3] O. Löhlein and M. Fritzsche, "Classification of GPR data for mine detection based on hidden Markov models," in *Proc. EUREL Conf. Detect. Abandon. Landmines*, Edinburgh, U.K., Oct. 1998, pp. 96–100.
- [4] C. Yang, "Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 743–753, May 2005.

- [5] X. Miao, M. Azimi-Sadjadi, B. Tian, A. Dubey, and N. Witherspoon, "Detection of mines and minelike targets using principal component and neural methods," *IEEE Trans. Neural Netw.*, vol. 9, no. 3, pp. 454–463, May 1998.
- [6] B. Baertlein, "Neural network based multi-sensor fusion for mine detection," Ballena Systems Corporation, Los Alamos, NM, Tech. Rep. N61331-93-C-0050, 1994.
- [7] S. Sheedvash and M. Azimi-Sadjadi, "Structural adaptation in neural networks with applications to land mine detection," in *Proc. IEEE Int. Conf. Neural Netw.*, 1997, pp. 1443–1447.
- [8] J. Zhang, Q. Liu, and B. Nath, "Landmine feature extraction and classification of GPR data based on SVM method," in *Proc. Int. Symp. Neural Netw., Part I*, Aug. 2004, pp. 636–641.
- [9] P. Torrione and L. Collins, "Application of texture feature classification methods to landmine/clutter discrimination in off-road GPR data," in *Proc. IGARSS*, 2004, pp. 1621–1624.
- [10] H. Frigui, P. Gader, and K. Satyanarayana, "Landmine detection with ground penetrating radar using fuzzy k-nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2004, vol. 3, pp. 1745–1749.
- [11] P. Gader, J. Keller, H. Frigui, H. Liu, and D. Wang, "Landmine detection using fuzzy sets with GPR images," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, May 1998, vol. 1, pp. 232–236.
- [12] H. Frigui, K. Satyanarayana, and P. Gader, "Detection of land mines using fuzzy and possibilistic membership functions," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, May 2003, vol. 2, pp. 834–839.
- [13] P. Gader, W. Lee, and A. Mendez-Vasquez, "Continuous choquet integrals with respect to random sets with applications to landmine detection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2004, vol. 1, pp. 523–528.
- [14] P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, and A. Zare, "Multi-sensor and algorithm fusion with the choquet integral: Applications to landmine detection," in *Proc. Geosci. and Remote Sens. Symp.*, 2004, vol. 3, pp. 1605–1608.
- [15] L. Yan, R. Dodier, M. Mozer, and R. Wolniewicz, "Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic," in *Proc. 20th Int. Conf. Mach. Learn.*, Washington, DC, 2003, pp. 848–855.
- [16] A. Rakotomamonjy, "Optimizing area under ROC curve with SVMs," in *Proc. ROCAI*, 2004, pp. 71–80.
- [17] P. D. Gader, R. Grandhi, W.-H. Lee, and J. N. Wilson, "Integration of ordered weighted averaging operators with feedforward neural networks for feature subset selection and pattern classification," *IEEE Trans. Syst., Man, Cybern.* submitted for.
- [18] L. Collins, P. Torrione, V. Munshi, C. Throckmorton, Q. Zhu, J. Clodfelter, and S. Frasier, "Algorithms for land mine detection using the NIITEK ground penetrating radar," in *Proc. SPIE Int. Soc. Opt.*, 2002, pp. 709–718.
- [19] P. Gader, W. Lee, and J. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.
- [20] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, Jan./Feb. 1988.
- [21] J. Fodor, J. Marichal, and M. Roubens, "Characterization of the ordered weighted averaging operators," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 2, pp. 236–240, May 1995.
- [22] R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators, Theory, and Applications*. Boston, MA: Kluwer, 1997.

Wen-Hsiung Lee received the B.S. degree in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1990, the Master and Ph.D. degrees from the University of Florida, Gainesville, in 1994 and 2001, respectively.

From 2002 to 2006, he was a Postdoctoral Researcher with the Department of Computer and Information Science and Engineering, University of Florida. Since August 2006, he has been a Senior Scientist with NIITEK Inc., Sterling, VA.



Paul D. Gader (M'87–SM'99) received the Ph.D. degree in mathematics for parallel image processing and applied mathematics research from the University of Florida, Gainesville, in 1986.

He has worked as a Senior Research Scientist with Honeywell's Systems and Research Center, as a Research Engineer and Manager with the Environmental Research Institute of Michigan, and as a faculty member with the University of Wisconsin—Oshkosh, the University of Missouri—Columbia, and the University of Florida, Gainesville,

where he is currently a Professor with the Department of Computer and Information Science and Engineering. He has been active in landmine-detection algorithm research since 1996. He has led teams that devised and tested several real-time algorithms in the field for mine detection. He has over 150 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including over 50 refereed journal articles.

Joseph N. Wilson received the B.S. degree in applied mathematics with the emphasis on computer science from Florida State University, Tallahassee, in 1977, and the M.S. degree in applied mathematics and computer science and the Ph.D. degree in computer science from the University of Virginia, Charlottesville, in 1980 and 1985, respectively.

Since 1984, he has been with the faculty of the Computer and Information Science and Engineering Department at the University of Florida, Gainesville, where he served as an Associate Chair from 1994 to 2001. His research interests include computer vision, image processing, programming languages, and file systems.

Minimum Classification Error Training for Choquet Integrals With Applications to Landmine Detection

Andres Mendez-Vazquez, Paul Gader, *Senior Member, IEEE*, James M. Keller, *Fellow, IEEE*, and Kenneth Chamberlin

Abstract—A novel algorithm for discriminative training of Choquet-integral-based fusion operators is described. Fusion is performed by Choquet integration of classifier outputs with respect to fuzzy measures. The fusion operators are determined by the parameters of fuzzy measures. These parameters are found by minimizing a minimum classification error (MCE) objective function. The minimization is performed with respect to a special class of measures, the Sugeno λ -measures. An analytic expression is derived for the gradient of the Choquet integral with respect to the Sugeno λ -measure. The new algorithm is applied to a landmine detection problem, and compared to previous techniques.

Index Terms—Choquet integral, fuzzy measures, least squared error (LSE), minimum classification error (MCE), Sugeno λ -measure.

I. INTRODUCTION

THE Choquet integral has been proposed as an aggregation operator for information fusion and pattern classification [1]–[12]. The application of fuzzy integrals to information fusion was first proposed by Tahani and Keller [13], [37]. Grabisch *et al.* [14], [1] proposed a least squares error methodology for training measures for the Choquet integral using quadratic programming and a heuristic gradient-descent algorithm. These approaches both suffer from two problems. They are sensitive to the values chosen for desired outputs and the number of parameters grows exponentially as a function of the number of information sources. In addition, for the gradient-descent method, heuristics must be used to insure that the monotonicity constraints of fuzzy measures are maintained. Chiang [15] proposed a method for using gradient descent to optimize Choquet integrals with respect to Sugeno λ -measures, but the formulas for the derivatives were incorrect [16].

This paper makes two novel contributions. First, an analytic expression is given for the derivative of the discrete Choquet integral with respect to a Sugeno λ -measure. Second, a minimum

classification error (MCE) approach for training the Choquet integral that uses the analytic derivation is developed. This new training approach reduces the number of parameters, removes the need to set desired outputs, and does not require heuristics to maintain the monotonicity constraints.

This paper is divided into the following sections. The first section deals with the basic definitions of fuzzy measure, Sugeno λ -measure, and discrete Choquet integral. After this preliminary overview, we review the training algorithm proposed by Grabisch [17]. Following this, we derive the gradient of the discrete Choquet integral with respect to Sugeno λ -measures. This gradient is then used to derive two different updating equations for gradient-descent-based optimization, the least squared error (LSE) and MCE, objective functions involving Choquet integrals and Sugeno λ -measures.

In the experimental section, we present information fusion and classification results. The classification results are obtained by applying the MCE methodology to Choquet integrals on standard data sets and compare them to those given in [8]. The results show that the methodology can be used to train classifiers for multiple classes and can perform as well or better than existing methods. Information fusion results are obtained using the MCE algorithm in the context of landmine detection. We compare them to results obtained using LSE with respect to Sugeno and general measures in the context of landmine detection. The results point to improvement of this fusion over the individual detectors, and the classic LSE training algorithm using one or multiple measures.

II. FUZZY MEASURES AND THE CHOQUET INTEGRAL

We first define some of the basic concepts behind the theory of fuzzy measures. These definitions can be found in [18], [14], [19], and [20].

Definition 1: Let $X = \{x_1, \dots, x_n\}$ be any finite set. A discrete **fuzzy measure** on X is a function $\mu : 2^X \rightarrow [0, 1]$ with the following properties:

- 1) $\mu(\emptyset) = 0$ and $\mu(X) = 1$;
- 2) given $A, B \in 2^X$, if $A \subset B$ then $\mu(A) \leq \mu(B)$ (monotonicity property).

For our purposes, the set X is considered to contain the names of sources of information (features, algorithms, agents, features, sensors, etc.), and for a subset $A \subseteq X$, $\mu(A)$ is considered to be the *worth* of this subset of information.

The Sugeno λ -measures are a special class of fuzzy measures. In keeping with notational convention, we refer to this class of measures using g instead of μ .

Manuscript received January 9, 2006; revised July 6, 2006 and October 24, 2006. This work was supported in part by the U.S. Army Research Office and U.S. Army Research Laboratory under Grant W911NF-05-1-0067 and was accomplished under Cooperative Agreement Number DAAD19-02-2-0012.

A. Mendez-Vazquez and P. Gader are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611-6120 USA (e-mail: amendez@cise.ufl.edu; pgader@cise.ufl.edu).

J. M. Keller is with the Department of Electrical Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211-2060 USA (e-mail: kellerj@missouri.edu).

K. Chamberlin was with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611-6120 USA. He is now with the CAE USA, Tampa, FL 33634-2481 USA.

Digital Object Identifier 10.1109/TFUZZ.2007.902024

Definition 2: Let $X = \{x_1, \dots, x_n\}$ be any finite set and let $\lambda \in (-1, +\infty)$. A **Sugeno λ -measure** is a function g from 2^X to $[0, 1]$ with the following properties:

- 1) $g(X) = 1$;
- 2) if $A, B \subseteq X$ with $A \cap B = \emptyset$, then

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B). \quad (1)$$

It can be shown that a set function satisfying the conditions in Definition 2 is a fuzzy measure. In particular, equation (1) implicitly imposes the monotonicity constraints on the Sugeno measures. As a convention, the measure of a singleton set $\{x_i\}$ is called a density and is denoted by $g_i = g(\{x_i\})$. In addition, we have that λ satisfies the property

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g_i). \quad (2)$$

The parameter λ is specific to this class of measures and can be computed from (2) once the densities are known. Tahani and Keller showed that this polynomial has a real root greater than -1 and several researchers have observed that this polynomial equation is easily solved numerically [21], [13], [37], [12]. By property (1), specifying a Sugeno λ -measure on a set X with n elements only requires specifying the n different densities, thereby reducing the number of free parameters from $2^n - 2$ to n .

To fuse evidence supplied by different sources of information from a discrete fuzzy set of X , we use the discrete **Choquet integral**.

Definition 3: Let f be a function from $X = \{x_1, \dots, x_n\}$ to $[0, 1]$. Let $\{x_{\sigma(1)}, \dots, x_{\sigma(n)}\}$ denote a reordering of the set X such that $0 \leq f(x_{\sigma(1)}) \leq \dots \leq f(x_{\sigma(n)})$, and let $A_{(i)}$ be a collection of subsets defined by $A_{(i)} = \{x_{\sigma(i)}, \dots, x_{\sigma(n)}\}$. Then, the discrete **Choquet integral** of f with respect to a fuzzy measure μ on X is defined as

$$\begin{aligned} C_\mu(f) &= \sum_{i=1}^n \mu(A_{(i)}) (f(x_{(i)}) - f(x_{(i-1)})) \\ &= \sum_{i=1}^n f(x_{(i)}) (\mu(A_{(i)}) - \mu(A_{(i+1)})) \end{aligned} \quad (3)$$

where we take $f(x_{(0)}) \equiv 0$, $A_{(n+1)} \equiv \emptyset$, and $x_{(i)} \equiv x_{\sigma(i)}$.

The function f is a particular instance of the partial support (evidence) supplied by each source of information in determining the confidence in an underlying hypothesis. The integral fuses this objective support with the worth (averagibility) of various subsets of the information sources. We remark that in the general definition of a Choquet integral, the function f does not need to have range $[0, 1]$. Our methodology relies on using histograms of the data, and therefore, naturally normalizes the function values to the range $[0, 1]$ using (8) and (7).

Some extra notation is needed to make a reference to objects in the classification problem. Let Ω denote the set of objects to be classified. Each information source x_i for $i = 1, \dots, n$ is a function $x_i : \Omega \rightarrow [0, 1]$. For each $\omega \in \Omega$, we define $f_\omega : X \rightarrow [0, 1]$ by $f_\omega(x_i) = x_i(\omega)$.

We now describe two previously published methods for learning fuzzy measures by minimizing LSE cost functions.

III. LSE COST FUNCTIONS FOR FUZZY MEASURES

One of the first **cost functions** used to learn the values for the discrete measure was proposed by Grabisch *et al.* [17], [14], [22]. Given classes C_1, \dots, C_n , they proposed a mean squared error (MSE) criterion, where the difference desired outputs α_i for $i = 1, \dots, n$ and the actual outputs $C_\mu(f)$ are minimized under constraints. The cost function is

$$E^2 = \sum_{\omega \in C_1} (C_\mu(f_\omega) - \alpha_1)^2 + \dots + \sum_{\omega \in C_n} (C_\mu(f_\omega) - \alpha_n)^2. \quad (4)$$

This cost function can be reduced to a quadratic optimization subject to linear constraints, i.e.,

$$\min \frac{1}{2} u^t D u + \Gamma u + \alpha, \quad \text{s.t. } A u + b \leq 0. \quad (5)$$

In particular D, Γ, A , and α^1 are determined by the data and what outputs need to be learned for each respective class.

An immediate problem in this approach is the use of the same measure for the different classes. Grabisch and Nicolas [22] addressed this problem with a modified version of (4) for a two-class problem

$$\begin{aligned} E^2 &= \sum_{\omega \in C_1} (C_{\mu_1}(\phi_1[f_\omega]) - C_{\mu_2}(\phi_2[f_\omega]) - \alpha_1)^2 + \dots \\ &\quad + \sum_{\omega \in C_2} (C_{\mu_2}(\phi_2[f_\omega]) - C_{\mu_1}(\phi_1[f_\omega]) - \alpha_n)^2 \end{aligned} \quad (6)$$

where ϕ_1 and ϕ_2 are functions that compute class specific confidence values from the information source outputs. For example, in information fusion, we use

$$\begin{aligned} \phi_i(f_\omega(x)) &= p(y \leq x(\omega) | C_1)^{2-i} (1 - p(y \leq x(\omega) | C_2))^{i-1} \\ &= F_{C_1}(x | \omega)^{2-i} (1 - F_{C_2}(x | \omega))^{i-1}, \quad \text{for } i = 1, 2 \end{aligned} \quad (7)$$

whereas in classification, we use

$$\phi_i(f_\omega(x)) = P(x(\omega) | C_i), \quad \text{for } i = 1, \dots, n. \quad (8)$$

Note that in (7), $\phi_i(f_\omega(x)) \in [0, 1]$. In (8), the values of x are generally quantized so the distribution is discrete and $\phi_i(f_\omega(x)) \in [0, 1]$. We can employ a similar procedure as the one in (4) to convert this cost function (6) into a quadratic problem under linear constraints. Two problems are encountered when solving these quadratic programs. First, specifying a general fuzzy measure requires specification of $2^n - 2$ parameters, which is clearly exponential. Second, the solution can be sensitive to the desired outputs. We therefore consider fuzzy measures with n free parameters and a cost function that avoids the use of desired outputs.

IV. GRADIENT OF THE DISCRETE CHOQUET INTEGRAL WITH RESPECT TO A SUGENO λ -MEASURE

It is desirable to implicitly, rather than explicitly as in (5), maintain the fuzzy measure constraint. This would allow the adoption of a simple gradient-descent scheme for optimization. A measure that implicitly enforces the fuzzy constraint is the

¹The α s can be interpreted as the ideal result of evaluating a function F' , the function to be approximated, in an input x .

Sugeno λ -measure. That is, the constraint that $g(A) \leq g(B)$ if $A \subseteq B$ is always satisfied from property (2) of Definition 1. Thus, it is desirable to obtain an expression for the gradient of the Choquet integral with respect to the λ -measure.

The gradient is obtained by differentiating the discrete Choquet integral (3) with respect to the densities of the Sugeno λ -measure. Thus, each partial derivative of $C_g(f_\omega)$ with respect to g_j ² is given by

$$\frac{\partial C_g(f_\omega)}{\partial g_j} = \sum_{i=1}^n \frac{\partial g(A_{(i)})}{\partial g_j} (f_\omega(x_{(i)}) - f_\omega(x_{(i-1)})) \quad (9)$$

To derive $(\partial g(A_{(i)}))/(\partial g_j)$, consider that according to (1)

$$g(A_{(i)}) = g(\{x_{(i)}\} \cup A_{(i+1)}) = g_{(i)} + g(A_{(i+1)}) + \lambda g_{(i)} g(A_{(i+1)}) \quad (10)$$

The partial derivative of this last (10) with respect to a density g_j is equal to

$$\begin{aligned} \frac{\partial g(A_{(i)})}{\partial g_j} &= \frac{\partial g_{(i)}}{\partial g_j} + \frac{\partial g(A_{(i+1)})}{\partial g_j} \\ &\quad + \frac{\partial \lambda}{\partial g_j} g_{(i)} g(A_{(i+1)}) + \dots \\ &\quad + \lambda \frac{\partial g_{(i)}}{\partial g_j} g(A_{(i+1)}) + \lambda g_{(i)} \frac{\partial g(A_{(i+1)})}{\partial g_j} \end{aligned} \quad (11)$$

Several cases need to be considered to obtain a general rule for this derivative (see Appendix I). However, we still need to derive an expression for $(\partial \lambda)/(\partial g_j)$. First, a unique λ for a given set of densities can be found by solving (2). After some work, we finish with the following term (see Appendix I):

$$\frac{\partial \lambda}{\partial g_j} = \frac{\lambda^2 + \lambda}{(1 + g_j \lambda) \left[1 - (\lambda + 1) \sum_{i=1}^n \left(\frac{g_i}{1 + g_i \lambda} \right) \right]}, \quad \lambda \neq 0. \quad (12)$$

Then, we can use (10)–(12) to design a gradient-descent algorithm for any cost function involving the Choquet integral and the Sugeno λ -measure. We note that the constraints that the densities must lie in the interval $[0, 1]$ must be enforced. This is a standard practice in MCE applications such as maintain stochastic constraints in hidden Markov models [23]–[25]. For this, we employed the techniques of clipping and auxiliary variables. Specifically, in the latter case, we take

$$g_j = \frac{1}{1 + e^{-z_j}}, \quad \text{where } z_j \in (-\infty, \infty) \quad (13)$$

which has a well-known, well-behaved derivative. Since the densities are forced to lie between $[0, 1]$, the measures are guaranteed to be monotonic.

We have two immediate advantages of using this methodology. First, property (1) of Sugeno λ -measure preserves the fuzzy measures constraint, as long as the densities stay in the interval $[0, 1]$ during the gradient-descent iterations. Second, we obtain a reduction in the computational complexity because we

only need to calculate the changes in the densities of the Sugeno λ -measure.

V. LSE AND MSE

Two different cost functions, LSE and MCE, were considered.

A. Least Squared Error

The first cost function proposed under the Sugeno derivation is an LSE minimization for a two-class problem

$$E^2 = \frac{1}{2} \sum_{\omega \in C_1} (C_g(f_\omega) - \alpha_1)^2 + \frac{1}{2} \sum_{\omega \in C_2} (C_g(f_\omega) - \alpha_2)^2 \quad (14)$$

where the problem is defined in terms of a unique Sugeno λ -measure and a pair of desired outputs. Taking a partial derivative with respect to each of the densities g_j , we have

$$\begin{aligned} \frac{\partial E^2}{\partial g_j} &= \sum_{\omega \in C_1} (C_g(f_\omega) - \alpha_1) \frac{\partial C_g(f_\omega)}{\partial g_j} + \dots \\ &\quad + \sum_{\omega \in C_2} (C_g(f_\omega) - \alpha_2) \frac{\partial C_g(f_\omega)}{\partial g_j} \end{aligned} \quad (15)$$

Using this last equation together with the expression for the Sugeno λ -measure derivatives, it is possible to define a gradient-descent algorithm.

Although, the cost function (14) reduces the computational complexity, it is still dependant on desired outputs. This is a serious drawback of the LSE cost functions.

B. Minimum Classification Error

In MCE training [23]–[25], we do not consider cost functions that use desired outputs. We instead consider a cost function that depends on a difference between confidences of different classes [26], [27]. These differences are called **dissimilarity measures**. Note that for correct classification, dissimilarity measures are negative. The dissimilarity measure we use is

$$d_i(f_\omega) = -C_{\mu_i}(\phi_i[f_\omega]) + \max_{j, j \neq i} C_{\mu_j}(\phi_j[f_\omega]). \quad (16)$$

Note that (16) allows for multiple classes. The MCE algorithm requires differentiation. Note that the function max is differentiable almost everywhere with a very simple derivative given by

$$\begin{aligned} \frac{\partial \max(f(x_1), f(x_2), \dots, f(x_n))}{\partial x_i} &= \begin{cases} \frac{\partial f(x_i)}{\partial x_i}, & \text{if } f(x_i) = \max(f(x_1), f(x_2), \dots, f(x_n)) \\ 0, & \text{else} \end{cases} \end{aligned} \quad (17)$$

In MCE, we introduce a loss function. Some examples of loss function are

$$l_i(f_\omega) = \begin{cases} (d_i(f_\omega))^\alpha, & \text{if } d_i(f_\omega) > 0 \\ 0, & \text{if } d_i(f_\omega) \leq 0 \end{cases}, \quad \alpha > 0, \alpha \rightarrow 0 \quad (18)$$

$$l_i(f_\omega) = \frac{1}{1 + e^{(-\alpha d_i(f_\omega))}}, \quad \alpha > 0. \quad (19)$$

²Note that λ is also a function of g_j . This can be seen in (2).

In our specific optimization, we combine (18) and (19) in a single loss function for information fusion

$$l_i(f_\omega) = \begin{cases} \frac{1}{1 + e^{(-\alpha d_i(f_\omega))}}, & d_i(f_\omega) > 0 \\ 0, & d_i(f_\omega) \leq 0 \end{cases} \quad (20)$$

For classification, we use a slightly modified version of the loss function (20)

$$l_i(f_\omega) = \begin{cases} 2 \left(\frac{1}{1 + e^{(-\alpha d_i(f_\omega))}} - \frac{1}{2} \right), & d_i(f_\omega) > 0 \\ 0, & d_i(f_\omega) \leq 0 \end{cases} \quad (21)$$

They have the property that correctly classified samples have zero loss. Thus, only samples that are not correctly classified are taken in consideration for the accumulative change in the optimization.

With this loss function (20) and the dissimilarity measure (16), we have the following cost function for n classes:

$$E = \sum_{\omega \in C_1} l_1(f_\omega) + \dots + \sum_{\omega \in C_n} l_n(f_\omega). \quad (22)$$

Hence, for the loss function (20)

$$\begin{aligned} \frac{\partial E}{\partial g_i^j} = & \sum_{\omega \in C_1} l_1(f_\omega)(1 - l_1(f_\omega)) \frac{\partial d_1(f_\omega)}{\partial g_i^j} + \dots \\ & + \sum_{\omega \in C_n} l_n(f_\omega)(1 - l_n(f_\omega)) \frac{\partial d_n(f_\omega)}{\partial g_i^j} \end{aligned} \quad (23)$$

where g_i^j represents the j th density for i th class. Now, the term $\frac{\partial d_k(f_\omega)}{\partial g_i^j}$ is equal to

$$\frac{\partial d_k(f_\omega)}{\partial g_i^j} = \begin{cases} -\frac{\partial C_{g_k}(f_\omega)}{\partial g_i^j}, & \text{if } k = i \\ \frac{\partial C_{g_i}(f_\omega)}{\partial g_i^j}, & \text{if } k \neq i \text{ and } C_{g_i}(f_\omega) = \max_{s \neq k} \{C_{g_s}(f_\omega)\} \\ 0, & \text{if } k \neq i \text{ and } C_{g_i}(f_\omega) \neq \max_{s \neq k} \{C_{g_s}(f_\omega)\}. \end{cases} \quad (24)$$

The derivations for the loss function (21) can be obtained in the same way.

We can use then (9)–(12) to obtain a gradient-descent algorithm for the MCE cost function (22).

This new optimization has the advantages that we have been looking for. First, each class is represented by a unique measure, and second, no desired outputs are necessary whatsoever.

VI. EXPERIMENTS

The LSE and MCE training methods were applied to a two-class algorithm fusion problem in landmine detection and to some standard data sets for pattern classification. We first discuss the fusion experiments and then the classification experiments.

The landmine detection problem involved processing ground penetrating radar (GPR) sensor returns. It is well described in

the literature but is briefly specified here. The goal is to discriminate between regions of ground that contain buried landmines from regions of ground that do not contain buried landmines. GPR measurements were made at multiple locations, some of which contain landmines and some of which do not. Multiple detection algorithms have been developed by numerous researchers to process samples obtained from these sensors, as described in [28]–[34]. Each detection algorithm involves a complex sequence of processes including signal processing, feature extraction, and classification. The algorithms produce confidence values as output. The larger the confidence value, the more likely it is that the input sample was acquired over a region of ground containing a landmine.

The data set contained 2422 8-D samples, each containing one confidence value from each of the eight detection algorithms used in the detection problem. The data set contained 271 mines samples and 2151 nonmine samples.

Three different information fusion algorithms were considered: LSE for a general measures, LSE for Sugeno λ -measures, and MCE for Sugeno λ -measures.

The probability of detection (PD) and the probability of false alarm (PFA) are used as performance measures. They are defined as follows:

$$PD(t) = \frac{|\{\omega \in \text{Mines} : C_{\mu_{\text{Mines}}}(f_\omega) \geq t\}|}{|\{\omega \in \text{Mines}\}|} \quad (25)$$

$$PFA(t) = \frac{|\{\omega \in \text{Nonmines} : C_{\mu_{\text{Nonmines}}}(f_\omega) \geq t\}|}{|\{\omega \in \text{Nonmines}\}|} \quad (26)$$

where $|\cdot|$ denotes the set cardinality.

Since gradient descent is sensitive to initialization, we run N -fold cross-validation M times to obtain a realistic estimate of the expected performance (for one experiment, $N = 5$ and $M = 20$). In addition, since LSE performance depends on the choice of desired outputs and the results are sensitive to this choice, we average over a range of reasonable desired outputs. The following pseudocode depicts the experimental procedure. In this pseudocode, weights refers to the parameters of the measure to be learned. The function compute receiver operating characteristic (ROC) computes for us the (25) and (26) for all t in the range of detections. The values α_1 and α_2 represent the desired outputs of the fuzzy integral in the range $[0, 1]$, for mines and nonmines, respectively. The value selected for α_1 ranges between 0.5 and 1 for mines, and for α_2 , we choose values between 0.0 and $\alpha_1 - 0.1$. This values are used because we want higher desired outputs for mines and lower outputs for nonmines. Each fold in the cross-validation scheme is represented by A_i .

General Testing Algorithm

- Initialize
 - 1) Data set = $\{A_i\}_{i=1}^N$ where $A_i \cap A_j = \emptyset$, if $i \neq j$.
 - 2) Number of repetitions for experiment = M .
- For $i = 1$ to M do
 - 1) randomly initialize weights
 - 2) $K = 1$

- 3) for $j = 1$ to N
 - if algorithm is LSE (this varies the desired output)
 - * for $\alpha_1 = 0.5$ to $1(\alpha_1, \alpha_2)$
 - for $\alpha_2 = 0.0$ to $\alpha_1 - 0.1$
 - $C_{ijK} = \text{Train}(\text{Weights}, \bigcup_{i=1, i \neq j}^N \{A_i\}, \alpha_1, \alpha_2)$
 - $K = K + 1$
 - else
 - * $C_{ij} = \text{Train}(\text{Weights}, \bigcup_{i=1, i \neq j}^N \{A_i\})$.
 - if algorithm is LSE (this varies the desired output)
 - for $i = 1$ to M
 - * for $j = 1$ to $K - 1$
 - $(\text{PD}_{ij}, \text{PFA}_{ij}) = \text{computeROC}(\{C_{i1j}, \dots, C_{iNj}\})$
 - else
 - for $i = 1$ to M
 - * $(\text{PD}_i, \text{PFA}_i) = \text{computeROC}(\{C_{i1}, \dots, C_{iN}\})$
 - if algorithm is LSE (this varies the desired output)
 - $\text{PD} = \frac{1}{M} \sum_{i=1}^M \left(\frac{1}{K-1} \sum_{j=1}^{K-1} \text{PD}_{ij} \right)$
 - $\text{PFA} = \frac{1}{M} \sum_{i=1}^M \left(\frac{1}{K-1} \sum_{j=1}^{K-1} \text{PFA}_{ij} \right)$
 - else
 - $\text{PD} = \frac{1}{M} \sum_{i=1}^M \text{PD}_i$
 - $\text{PFA} = \frac{1}{M} \sum_{i=1}^M \text{PFA}_i$.

Here, PD_{ijK} and PFA_{ijK} represent the PD and PFA of the j th cross-validation fold of the i th experiment and the K th variations in desired outputs for the LSE training functions. In a similar fashion, PD_{ij} and PFA_{ij} of the j th cross-validation fold of the i th experiment for the MCE training function.

Before examining the results from each algorithm, we show the sensitivity of the LSE training for the two measures used in the experiments. The ROC plots in Figs. 1 and 2 show some of the variations in the PD and PFA due to random initialization under the different desired outputs in a single experiment. We can see that different desired outputs produce different ROC curves. In addition, the best ROC curve is not obtained using ideal values like zero for nonmines and one for mines, but non-intuitive values of 0.8 for mines and 0.2 for nonmines in the case of a Sugeno λ -measure, and 0.5 for mines and 0.1 for nonmines in the case of a general measure. These figures show the sensitivity of LSE schemes to desired outputs and random initialization.

Now, we can show the results obtained from each algorithm. In Table I, average Sugeno λ -measure trained via LSE is compared with each individual detector. For PDs ranging from 80%

to 100%, the table shows the PFA achieved by the Choquet integral with respect to Sugeno λ -measure, the PFA achieved by each detector, and the reduction of PFA achieved by the Choquet integral with respect to Sugeno λ -measure compared to each detector. The percentage of reduction ranges between 0.09% and 51.84%. Although a Choquet integral with respect to a Sugeno λ -measure trained with LSE performs better than many of the individual results, it is still worse than the best possible detectors (detectors 6 and 7).

In Table II, we compare individual detectors against the general measure trained using an LSE cost function. It is clear that general measures trained using LSE improve a certain amount over Sugeno λ -measures trained using LSE. This range of improvement is between 3.25% and 55.60%. However, the Choquet integral with respect to a general measure trained with LSE is still not better than the best detectors (detectors 6 and 7).

Table III shows that in contrast to the Sugeno λ -measure and the general measure trained with LSE, the Sugeno λ -measure trained with MCE is, in general, better than all the individual detectors, with a range of improvement between 0.44% and 65.07%.

Table IV shows the improvement of MCE over the LSE. The range of improvement is between 11.06% and 37.51% with respect to the LSE cost functions.

It is possible for the Sugeno λ -measure and the general measure trained with LSE to be as good as the one trained by MCE. For this to happen, it is necessary to have a set of correct desired outputs. It is clear that depending on initialization these desired outputs can change. This is a limitation for general measures and Sugeno λ -measures under LSE optimizations, and of course, an advantage of MCE training.

The MCE training was also applied to the iris and breast cancer data and compared to the results shown in [8] (note that the appendicitis data is no longer at the machine learning website). The iris data is a three-class problem whereas the breast cancer data is a two-class problem. As in [8], tenfold cross validation was performed. We report the average error rates achieved in Table V. The average error rate achieved on the iris data was 4% whereas the average error rate achieved on the breast cancer data was 22.7%, which compares favorably with the results in [8].

The computational complexity of the proposed training algorithm is not high. First, the number of free parameters is only n whereas the number of free parameters for a general measure is $2^n - 2$. The final time complexity in big O notation is

$$\text{MCE complexity} = O(Kn \log(n) + HM(n^3 + Kn^2)) \quad (27)$$

where H is the number of iterations in the main loop, K is the total number of training samples, and M is the number of classes (See Appendix II). In comparison, the sequential quadratic optimization, used to solve quadratic problems under constraints, would finish with an exponential time complexity.

VII. CONCLUSION

In this paper, we developed an MCE algorithm to train Choquet integrals for fusion, and tested the training algorithm

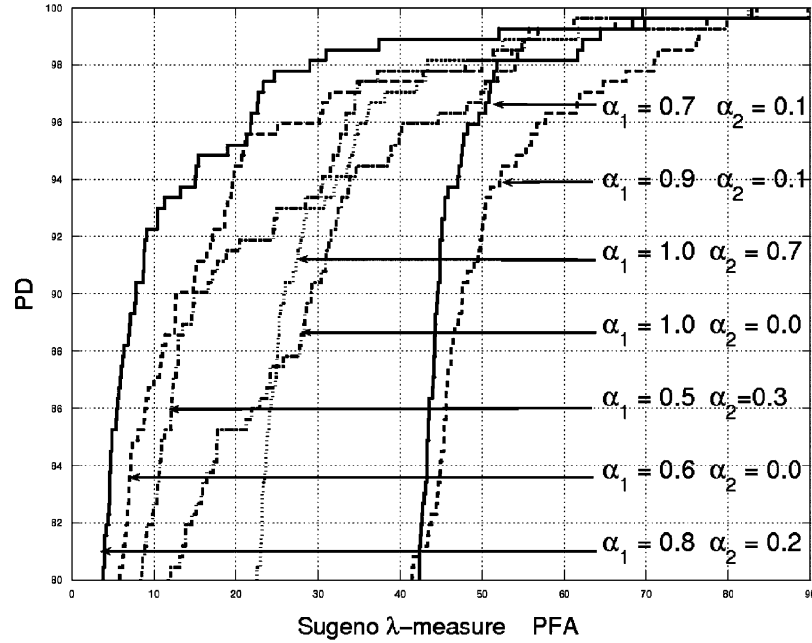


Fig. 1. Examples of MSE sensitivity to desired outputs for Sugeno λ -measure. α_1 and α_2 represent the desired outputs for mines and nonmines, respectively. The best ROC curve is obtained using 0.8 for mines and 0.2 for nonmines.

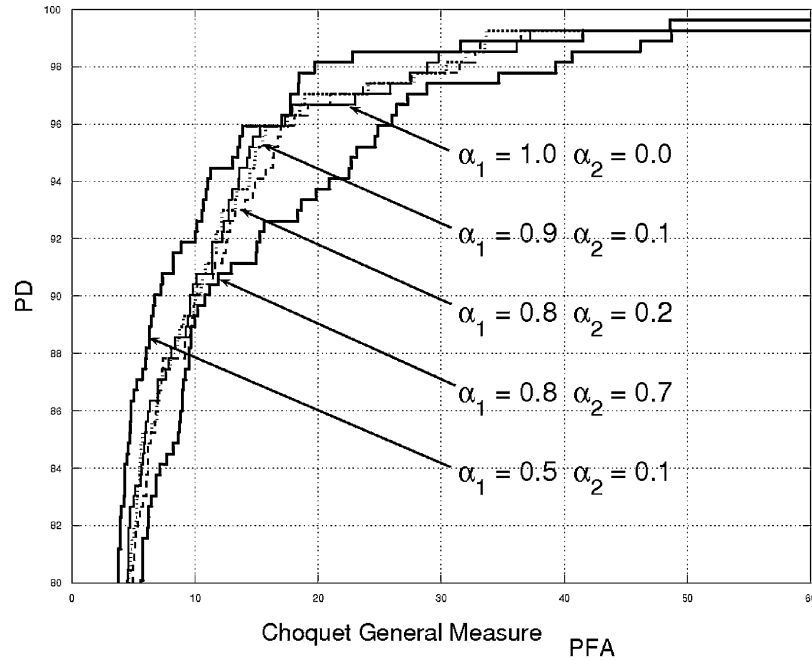


Fig. 2. Examples of squared error sensitivity to desired outputs for general measures. α_1 and α_2 represent the desired outputs for mines and nonmines respectively. The best ROC curve is obtained using 0.5 for mines and 0.1 for nonmines.

against the better known LSE training in a complex multi-classifier fusion data set from the application of landmine detection. The MCE approach allows training of Choquet integrals without requiring desired outputs. Although LSE training can do as well as MCE training, on average, LSE does significantly worse. In addition, we used the MCE algorithm to train pattern classifiers for standard data sets and the results compare favorably with existing results. The computational complexity of the proposed algorithm is low and the number of

free parameters grows only linearly with the number of inputs rather than exponentially.

A consequence of the exponential nature of the full measure is that any attempt to learn would require a new way to calculate it. For example, we could use Monte Carlo methods, which are extreme for solving high-dimensionality problems, to learn these measures, but you still have an exponential number of variables. Some thoughts have been given to this idea, but this is beyond the scope of this paper.

TABLE I
COMPARISON OF PFA FOR SUGENO λ -MEASURE TRAINED WITH LSE AGAINST DIFFERENT DETECTORS

PD	PFA λ -measure	PFA Det1	% Red.	PFA Det2	% Red.	PFA Det3	% Red.
100.00	98.28	98.37	0.09%	95.40	-3.02%	95.07	-3.37%
98.00	32.79	57.65	43.11%	68.11	51.85%	66.99	51.05%
96.00	24.21	30.03	19.38%	45.23	46.47%	40.31	39.93%
94.00	18.51	24.69	25.01%	31.06	40.39%	28.36	34.72%
92.00	13.56	19.15	29.20%	15.48	12.41%	18.83	27.98%
90.00	10.32	16.32	36.77%	13.67	24.51%	14.64	29.54%
88.00	8.70	14.27	39.02%	12.13	28.27%	12.18	28.54%
86.00	7.16	12.13	41.02%	10.04	28.73%	11.11	35.59%
84.00	5.96	10.18	41.44%	9.07	34.23%	9.34	36.19%
82.00	5.43	8.32	34.79%	7.62	28.82%	7.81	30.52%
80.00	4.79	7.39	35.20%	7.16	33.10%	6.65	27.95%

PD	PFA λ -measure	PFA Det4	% Red.	PFA Det5	% Red.	PFA Det6	% Red.
100.00	98.28	89.31	-10.05%	76.99	-27.66%	77.03	-27.58%
98.00	32.79	57.46	42.93%	49.09	33.20%	35.24	6.94%
96.00	24.21	35.80	32.36%	31.94	24.19%	22.97	-5.43%
94.00	18.51	27.24	32.05%	24.31	23.86%	13.16	-40.70%
92.00	13.56	16.36	17.14%	15.53	12.67%	10.93	-24.12%
90.00	10.32	12.13	14.96%	12.69	18.70%	9.44	-9.34%
88.00	8.70	9.72	10.42%	10.51	17.16%	6.79	-28.23%
86.00	7.16	8.69	17.68%	8.37	14.48%	5.25	-36.23%
84.00	5.96	8.00	25.43%	7.72	22.74%	5.21	-14.51%
82.00	5.43	6.65	18.37%	6.37	14.80%	4.60	-17.91%
80.00	4.79	6.00	20.13%	5.16	7.18%	4.60	-4.07%

PD	PFA λ -measure	PFA Det6	% Red.	PFA Det8	% Red.		
100.00	98.28	81.78	-20.18%	94.24	-4.29%		
98.00	32.79	38.73	15.32%	62.20	47.28%		
96.00	24.21	24.08	-0.54%	32.03	24.41%		
94.00	18.51	18.83	1.68%	26.69	30.63%		
92.00	13.56	13.58	0.11%	18.46	26.53%		
90.00	10.32	10.88	5.15%	15.20	32.12%		
88.00	8.70	8.32	-4.59%	13.44	35.22%		
86.00	7.16	6.69	-6.90%	12.27	41.69%		
84.00	5.96	5.63	-6.00%	10.55	43.50%		
82.00	5.43	5.30	-2.39%	9.67	43.88%		
80.00	4.79	4.14	-15.76%	9.25	48.23%		

TABLE II
COMPARISON OF PFA FOR GENERAL MEASURE TRAINED WITH LSE AGAINST DIFFERENT DETECTORS AT DIFFERENT THRESHOLDS

PD	PFA Gen. measure	PFA Det1	% Red.	PFA Det2	% Red.	PFA Det3	% Red.
100.00	91.17	98.37	7.33%	95.40	4.43%	95.07	4.11%
98.00	30.24	57.65	47.55%	68.11	55.61%	66.99	54.87%
96.00	21.28	30.03	29.13%	45.23	52.95%	40.31	47.19%
94.00	15.54	24.69	37.06%	31.06	49.97%	28.36	45.21%
92.00	12.91	19.15	32.61%	15.48	16.62%	18.83	31.45%
90.00	10.04	16.32	38.45%	13.67	26.52%	14.64	31.41%
88.00	8.43	14.27	40.93%	12.13	30.52%	12.18	30.78%
86.00	7.15	12.13	41.04%	10.04	28.76%	11.11	35.61%
84.00	6.06	10.18	40.48%	9.07	33.15%	9.34	35.15%
82.00	5.43	8.32	34.79%	7.62	28.82%	7.81	30.52%
80.00	4.94	7.39	33.22%	7.16	31.05%	6.65	25.75%

PD	PFA Gen. measure	PFA Det4	% Red.	PFA Det5	% Red.	PFA Det6	% Red.
100.00	91.17	89.31	-2.08%	76.99	-18.42%	77.03	-18.35%
98.00	30.24	57.46	47.38%	49.09	38.41%	35.24	14.20%
96.00	21.28	35.80	40.54%	31.94	33.36%	22.97	7.32%
94.00	15.54	27.24	42.97%	24.31	36.10%	13.16	-18.10%
92.00	12.91	16.36	21.12%	15.53	16.87%	10.93	-18.15%
90.00	10.04	12.13	17.22%	12.69	20.86%	9.44	-6.43%
88.00	8.43	9.72	13.23%	10.51	19.75%	6.79	-24.22%
86.00	7.15	8.69	17.71%	8.37	14.51%	5.25	-36.18%
84.00	6.06	8.00	24.21%	7.72	21.47%	5.21	-16.39%
82.00	5.43	6.65	18.37%	6.37	14.79%	4.60	-17.91%
80.00	4.94	6.00	17.69%	5.16	4.34%	4.60	-7.25%

PD	PFA Gen. measure	PFA Det7	% Red.	PFA Det8	% Red.		
100.00	91.17	81.78	-11.48%	94.24	3.26%		
98.00	30.24	38.73	21.92%	62.20	51.39%		
96.00	21.28	24.08	11.62%	32.03	33.55%		
94.00	15.54	18.83	17.48%	26.69	41.77%		
92.00	12.91	13.58	4.92%	18.46	30.06%		
90.00	10.04	10.88	7.67%	15.20	33.93%		
88.00	8.43	8.32	-1.32%	13.44	37.25%		
86.00	7.15	6.69	-6.87%	12.27	41.71%		
84.00	6.06	5.63	-7.73%	10.55	42.57%		
82.00	5.43	5.30	-2.40%	9.67	43.88%		
80.00	4.94	4.14	-19.30%	9.25	46.64%		

TABLE III
COMPARISON OF PFA IN MCE AGAINST DIFFERENT DETECTORS AT DIFFERENT THRESHOLDS

PD	PFA	PFA Det1	% Red.	PFA Det2	% Red.	PFA Det3	% Red.
100.00	94.65	98.37	3.78%	95.40	0.78%	95.07	0.44%
98.00	26.89	57.65	53.35%	68.11	60.52%	66.99	59.86%
96.00	15.80	30.03	47.40%	45.23	65.08%	40.31	60.81%
94.00	11.07	24.69	55.14%	31.06	64.34%	28.36	60.95%
92.00	8.07	19.15	57.89%	15.48	47.90%	18.83	57.16%
90.00	6.65	16.32	59.27%	13.67	51.38%	14.64	54.62%
88.00	5.45	14.27	61.81%	12.13	55.08%	12.18	55.25%
86.00	4.96	12.13	59.10%	10.04	50.58%	11.11	55.33%
84.00	4.38	10.18	57.01%	9.07	51.72%	9.34	53.16%
82.00	3.93	8.32	52.82%	7.62	48.51%	7.81	49.73%
80.00	3.61	7.39	51.10%	7.16	49.51%	6.65	45.63%
PD	PFA	PFA Det4	% Red.	PFA Det5	% Red.	PFA Det6	% Red.
100.00	94.65	89.31	-5.99%	76.99	-22.95%	77.03	-22.87%
98.00	26.89	57.46	53.20%	49.09	45.23%	35.24	23.69%
96.00	15.80	35.80	55.87%	31.94	50.54%	22.97	31.21%
94.00	11.07	27.24	59.35%	24.31	54.46%	13.16	15.83%
92.00	8.07	16.36	50.71%	15.53	48.05%	10.93	26.17%
90.00	6.65	12.13	45.23%	12.69	47.64%	9.44	29.58%
88.00	5.45	9.72	43.90%	10.51	48.12%	6.79	19.69%
86.00	4.96	8.69	42.91%	8.37	40.69%	5.25	5.53%
84.00	4.38	8.00	45.26%	7.72	43.28%	5.21	15.94%
82.00	3.93	6.65	40.94%	6.37	38.36%	4.60	14.70%
80.00	3.61	6.00	39.73%	5.16	29.95%	4.60	21.46%
PD	PFA	PFA Det6	% Red.	PFA Det8	% Red.		
100.00	94.65	81.78	-15.75%	94.24	-0.44%		
98.00	26.89	38.73	30.56%	62.20	56.77%		
96.00	15.80	24.08	34.40%	32.03	50.68%		
94.00	11.07	18.83	41.19%	26.69	58.50%		
92.00	8.07	13.58	40.58%	18.46	56.30%		
90.00	6.65	10.88	38.91%	15.20	56.28%		
88.00	5.45	8.32	34.50%	13.44	59.43%		
86.00	4.96	6.69	25.87%	12.27	59.56%		
84.00	4.38	5.63	22.19%	10.55	58.52%		
82.00	3.93	5.30	25.92%	9.67	59.40%		
80.00	3.61	4.14	12.64%	9.25	60.93%		

One can use formulas for the derivative of a Choquet integral with respect to the Sugeno λ -measure in any differentiable cost

function that includes Choquet integral. For example, we can optimize the fuzzy measures not against possible error outputs,

TABLE IV
MEAN MCE PFA AGAINST MEAN CENTRAL AND SUGENO PFA

PD	MCE PFA	PFA General Measure	% Red.	PFA Sugeno	% Red.
100.00	94.65	91.17	-3.82%	98.28	3.69%
98.00	26.89	30.24	11.07%	32.79	18.01%
96.00	15.80	21.28	25.78%	24.21	34.76%
94.00	11.07	15.54	28.73%	18.51	40.18%
92.00	8.07	12.91	37.51%	13.56	40.52%
90.00	6.65	10.04	33.83%	10.32	35.59%
88.00	5.45	8.43	35.35%	8.70	37.37%
86.00	4.96	7.15	30.63%	7.16	30.65%
84.00	4.38	6.06	27.77%	5.96	26.59%
82.00	3.93	5.43	27.66%	5.43	27.65%
80.00	3.61	4.94	26.77%	4.79	24.54%

TABLE V
COMPARISON OF MCE AGAINST SEVERAL OTHER CLASSIFIERS FOR IRIS AND BREAST CANCER DATA

Method	Iris Data(%)	Breast Cancer Data(%)
Linear	2.0	29
Quadratic	2.7	34.4
Nearest neighbor	4.0	34
Bayes independent	6.7	28.2
Bayes quadratic	16.0	34.4
Neuronal net	3.3	28.5
PVM rule	4.0	22.9
QUAD	3.3	31.5
CLMS	4.0	27.1
HLMS	4.7	22.6
WCIPP	4.0	26.2
MCE	4.0	22.73

but against the ROC curve itself. Thus, the derivation found here can be of value in other optimization methods.

APPENDIX I

DERIVATION OF THE DERIVATIVE OF THE CHOQUET INTEGRAL WITH RESPECT TO SUGENO λ -MEASURE

The gradient is obtained by differentiation the discrete Choquet integral

$$C_g(f) = \sum_{i=1}^n g(A_{(i)}) (f(x_{(i)}) - f(x_{(i-1)})) \quad (\text{A-1})$$

with respect to the densities of the Sugeno λ -measure. Thus, each partial derivative of $C_g(f)$ with respect to g_j is equal to

$$\frac{\partial C_g(f_\omega)}{\partial g_j} = \sum_{i=1}^n \frac{\partial g(A_{(i)})}{\partial g_j} (f_\omega(x_{(i)}) - f_\omega(x_{(i-1)})) \quad (\text{A-2})$$

To derive $(\partial g(A_{(i)}))/(\partial g_j)$, consider that, according to the property (1) of the Sugeno λ -measure

$$\begin{aligned} g(A_{(i)}) &= g(\{x_{(i)}\} \cup A_{(i+1)}) \\ &= g_{(i)} + g(A_{(i+1)}) + \lambda g_{(i)} g(A_{(i+1)}) \quad (\text{A-3}) \end{aligned}$$

It is well known that (A-4) can be derived from (A-4) assuming $\lambda \neq 0$

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g_i), \quad \lambda \neq 0. \quad (\text{A-4})$$

We can then consider the derivation of (A-3) for the case $\lambda \neq 0$. First, if $(i) = j$, we have that

$$\begin{aligned} \frac{\partial g(A_{(i)})}{\partial g_j} &= 1 + \frac{\partial g(A_{(i+1)})}{\partial g_j} + g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + \dots \\ &\quad + \lambda g(A_{(i+1)}) + \lambda g_{(i)} \frac{\partial g(A_{(i+1)})}{\partial g_j} \\ &= 1 + \lambda g(A_{(i+1)}) + g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + \dots \\ &\quad + (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j} \quad (\text{A-5}) \end{aligned}$$

by the multiplication rule for derivatives. In a similar way for $(i) \neq j$, we have that

$$\begin{aligned} \frac{\partial g(A_{(i)})}{\partial g_j} &= 0 + \frac{\partial g(A_{(i+1)})}{\partial g_j} + \frac{\partial \lambda}{\partial g_j} g_{(i)} g(A_{(i+1)}) + \dots \\ &\quad + \lambda g_{(i)} \frac{\partial g(A_{(i+1)})}{\partial g_j} \\ &= g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + \dots \\ &\quad + (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j} \quad (\text{A-6}) \end{aligned}$$

From (A-5) and (A-6) and the fact that $g(A_{(n+1)}) = 0$, we can obtain the following.

Case I) $(i) \neq n, (i) = j$

$$\begin{aligned} \frac{\partial g(A_{(i)})}{\partial g_j} &= 1 + \lambda g(A_{(i+1)}) + \dots \\ &\quad + g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + \dots \\ &\quad + (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j}. \end{aligned} \quad (\text{A-7})$$

Case II) $(i) \neq n, (i) \neq j$

$$\begin{aligned} \frac{\partial g(A_{(i)})}{\partial g_j} &= g_{(i)} g(A_{(i+1)}) \frac{\partial \lambda}{\partial g_j} + \dots \\ &\quad (1 + \lambda g_{(i)}) \frac{\partial g(A_{(i+1)})}{\partial g_j}. \end{aligned} \quad (\text{A-8})$$

Case III) $(i) = n, j = n$

$$\frac{\partial g(A_{(i)})}{\partial g_j} = 1. \quad (\text{A-9})$$

Case IV) $(i) = n, j \neq n$

$$\frac{\partial g(A_{(i)})}{\partial g_j} = 0. \quad (\text{A-10})$$

Now, we only need to obtain an expression for $(\partial \lambda)/(\partial g_j)$. Differentiating both sides of (A-4) with respect to g_j yields

$$\begin{aligned} \frac{\partial \lambda}{\partial g_j} &= \left(\lambda \prod_{i=1, i \neq j}^n (1 + \lambda g_i) \right) + \dots \\ &\quad + \left(\frac{\partial \lambda}{\partial g_j} \sum_{i=1}^n g_i \prod_{k=1, k \neq i}^n (1 + \lambda g_k) \right). \end{aligned} \quad (\text{A-11})$$

From this equation, we can get the following:

$$\begin{aligned} \frac{\partial \lambda}{\partial g_j} \left(1 - \sum_{i=1}^n g_i \prod_{k=1, k \neq i}^n (1 + \lambda g_k) \right) \\ = \left(\lambda \prod_{i=1, i \neq j}^n (1 + \lambda g_i) \right) \end{aligned} \quad (\text{A-12})$$

which can be reduced to

$$\frac{\partial \lambda}{\partial g_j} = \frac{\lambda \prod_{i=1, i \neq j}^n (1 + \lambda g_i)}{1 - \sum_{i=1}^n g_i \prod_{k=1, k \neq i}^n (1 + \lambda g_k)} \quad (\text{A-13})$$

and because we can rewrite $\lambda + 1 = \prod_{i=1}^n (1 + \lambda g_i)$ as $(1 + \lambda)/(1 + \lambda g_j) = \prod_{i=1, i \neq j}^n (1 + \lambda g_i)$, we have

$$\frac{\partial \lambda}{\partial g_j} = \frac{\lambda \frac{1 + \lambda}{1 + \lambda g_j}}{1 - \sum_{i=1}^n g_i \frac{1 + \lambda}{1 + \lambda g_i}}. \quad (\text{A-14})$$

We have finally that

$$\frac{\partial \lambda}{\partial g_j} = \frac{\lambda^2 + \lambda}{(1 + g_j \lambda) \left[1 - (\lambda + 1) \sum_{i=1}^n \left(\frac{g_i}{1 + g_i \lambda} \right) \right]}, \quad \lambda \neq 0. \quad (\text{A-15})$$

With (A-15) together with (A-5) and (A-6), we can get the derivative of the Choquet integral with respect to the Sugeno λ -measure for $\lambda \neq 0$.

Note that the derivation of (A-4) from (A-3) assumes that $\lambda \neq 0$ and that the resulting expression for $(\partial \lambda)/(\partial g_j)$ in (A-15) is undefined for $\lambda = 0$ (since $\sum_{i=1}^n g_i = 1$). We can apply L'Hopital's rule to see that $\lim_{\lambda \rightarrow 0} (\partial \lambda)/(\partial g_j) = n$. Hence, in the unlikely event that $\lambda = 0$ during training, one can take $(\partial \lambda)/(\partial g_j) = n$.

APPENDIX II

TIME-COMPLEXITY ANALYSIS OF THE MINIMUM CLASSIFICATION ERROR TRAINING ALGORITHM

For this analysis, we assume that $|X| = n$, there are M classes, and each has M_i elements. In addition, it is easy to prove that once the sorting is done for a sample, the calculation of the Choquet integral can be done in linear time for the Sugeno λ -measure. Then, calculating the Choquet integral with respect to the Sugeno λ -measure has asymptotic complexity $O(n \log(n) + n)$ [35]. In addition, calculating the roots for (2) has asymptotic complexity $O(n^3)$ [35], [36].

We present the pseudocode of the general algorithm with the order of operations of the computational complexity steps in parentheses:

General MCE Algorithm

- Set learning rate α
- for $i = 1$ to M
 - Sort all the samples of class C_i , $(O(M_i n \log(n)))$
- endfor
- Do
 - for $k = 1$ to M
 - * Set $\nabla E_k = 0$
 - * Calculate λ_k for each class C_k $(O(n^3))$
 - * Calculate for each density $g_j^{(k)}$ the partial derivative

$$\frac{\partial \lambda_k}{\partial g_j^{(k)}} = \frac{\lambda_k^2 + \lambda_k}{(1 + g_j^{(k)} \lambda_k) \left[1 - (\lambda_k + 1) \sum_{i=1}^n \left(\frac{g_i^{(k)}}{1 + g_i^{(k)} \lambda_k} \right) \right]}, \quad (O(n))$$

- * for $h = 1$ to M_i

- 1) Calculate $d_k(f_{\omega_h}) = -C_{g^{(k)}}(\phi_k[f_{\omega_h}]) + \max_{j,j \neq k} C_{g^{(j)}}(\phi_j[f_{\omega_h}])$, ($O(Mn)$)
- 2) Calculate $l_k(f_{\omega_h})$, ($O(1)$)
- 3) Calculate $g^{(k)}(A_{(i)})$ for all $i = 1, \dots, n$ ($O(n)$)
- 4) Calculate for all $i = 1, \dots, n$ the partial derivative of $g^{(k)}(A_{(i)})$ with respect $g_j^{(k)}$ for all $j = 1, \dots, n$

$$\begin{aligned} \frac{\partial g^{(k)}(A_{(i)})}{\partial g_j^{(k)}} &= \frac{\partial g_{(i)}^{(k)}}{\partial g_j^{(k)}} + \dots \\ &+ \frac{\partial g^{(k)}(A_{(i+1)})}{\partial g_j^{(k)}} + \dots \\ &+ \frac{\partial \lambda_k}{\partial g_j^{(k)}} g_{(i)}^{(k)} g^{(k)}(A_{(i+1)}) + \dots \\ &+ \lambda_k \frac{\partial g_{(i)}^{(k)}}{\partial g_j^{(k)}} g^{(k)}(A_{(i+1)}) + \dots \\ &+ \lambda_k g_{(i)}^{(k)} \frac{\partial g^{(k)}(A_{(i+1)})}{\partial g_j^{(k)}}, (O(1)) \end{aligned}$$

- 5) Calculate the partial derivative $C_{g^{(k)}}(f_{\omega_h})$ with respect to $g_j^{(k)}$ for all $j = 1, \dots, n$

$$\begin{aligned} \frac{\partial C_{g^{(k)}}(f_{\omega_h})}{\partial g_j^{(k)}} &= \sum_{i=1}^N \frac{\partial g^{(k)}(A_{(i)})}{\partial g_j^{(k)}} \\ &\quad * (f_{\omega}(x_{(i)}) - f_{\omega}(x_{(i-1)})), (O(n)) \end{aligned}$$

- 6) Calculate for each density $g_j^{(k)}$ the quantity $D_{jkh} = l_j(f_{\omega_h})(1 - l_j(f_{\omega_h}))(\partial d_j(f_{\omega_h})) / (\partial g_j^{(k)})(O(1))$.
- 7) For each $g_j^{(k)}$, set $\dot{g}_j^{(k)} = g_j^{(k)} - \alpha * D_{jkh}$, ($O(n)$)
- 8) $\nabla E_k = \nabla E_k + (D_{1kh}, \dots, D_{Nkh})^T$, ($O(1)$)

* endfor

– endfor

• while $\|(\nabla E_1, \dots, \nabla E_M)^T\| > \epsilon$

$O(KM) + \dots$	Time complexity of calculating all loss functions $l_k(f_{\omega_h})$.
$O(KMn) + \dots$	Time complexity of calculating all measures $g^{(k)}(A_{(i)})$.
$O(KMn^2) + \dots$	Time complexity of calculating all partial derivatives of $g^{(k)}$.
$O(KMn) + \dots$	Time complexity of calculating all partial derivatives of $C_{g^{(k)}}(f_{\omega_h})$.
$O(KMn) + \dots$	Time complexity of calculating all D_{jkh} .
$O(KMn) + \dots$	Time complexity of updating all $g_j^{(k)}$.
$O(KM) + \dots$	Time complexity for updating all ∇E_k .

We can rewrite this time complexity as

$$O(Kn \log(n) + Mn^3 + KMn^2). \quad (A-16)$$

Thus, we have that the time complexity for a single iteration in the MCE is

$$\begin{aligned} \text{Time complexity for MCE single iteration} \\ = O(Kn \log(n) + Mn^3 + KMn^2). \quad (A-17) \end{aligned}$$

Then, assuming H iterations in the main *while* loop, we obtain the time complexity for the MCE

$$\begin{aligned} \text{Time complexity for MCE} \\ = O(Kn \log(n) + HM(n^3 + Kn^2)). \quad (A-18) \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank X. Zhang for his assistance with the landmine dataset and Y.-C. Lin from the University of Missouri-Columbia who suggested an early version of the implicit derivative. They would also like to thank R. Harmon, R. Weaver, P. Howard, M. Cathcart, and W. Clarke for their support of this work and L. Collins and P. Torriane of Duke University and F. Clodfelter and others from NIITEK, Inc., for their technical discussions, insights, cooperation, and GPR data. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Office, U.S. Army Research Laboratory, or the U.S. Government. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

- [1] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. IEEE 4th Int. Conf. Fuzzy Syst.*, Yokohama, Japan, Mar. 1995, pp. 145–150.
- [2] D. Dubois, M. Grabisch, F. Modave, and H. Prade, "Relating decision under uncertainty and multicriteria decision making models," *Int. J. Intell. Syst.*, vol. 15, no. 10, pp. 967–979, 2000.
- [3] M. Grabisch, H. Nguyen, and E. Walker, *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*. Dordrecht, The Netherlands: Kluwer, 1995.
- [4] M. Grabisch, "Modelling data by the Choquet integral," in *Information Fusion in Data Mining*, V. Torra, Ed. Heidelberg, Germany: Springer-Verlag, 2003, pp. 135–148.

First, define $K = \sum_{i=1}^M M_i$ to be the total number of samples. Now, the time complexity for a single iteration is

$O(Kn \log(n)) + \dots$	Time complexity of sorting all the samples
$O(Mn^3) + \dots$	Time complexity of calculating $\lambda_1, \dots, \lambda_M$.
$O(Mn^2) + \dots$	Time complexity of calculating $\nabla \lambda_1, \dots, \nabla \lambda_M$.
$O(KMn) + \dots$	Time complexity of calculating all dissimilarity functions $d_k(f_{\omega_h})$.

- [5] C. Labreuche and M. Grabisch, "The Choquet integral for the aggregation of interval scales in multicriteria decision making," *Fuzzy Sets Syst.*, vol. 137, pp. 11–26, 2003.
- [6] J.-L. Marichal, "Aggregation of interacting criteria by means of the discrete Choquet integral," in *Aggregation Operators: New Trends and Applications*, ser. Studies in Fuzziness and Soft Computing, T. Calvo, G. Mayor, and R. Mesiar, Eds. Heidelberg, Germany: Springer-Verlag, 2002, vol. 97, pp. 224–244.
- [7] F. Modave and M. Grabisch, "Preference representation by a Choquet integral: Commensurability hypothesis," in *Proc. 7th Int. Conf. Inf. Process. Manag. Uncertainty Knowl.-Based Syst. (IPMU)*, Paris, France, Jul. 1998, pp. 164–171.
- [8] K. Xu, Z. Wang, P.-A. Heng, and K.-S. Leung, "Classification by non-linear integral projections," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 187–200, Apr. 2003.
- [9] S. Auephanwirayakul, J. Keller, and P. D. Gader, "Generalized Choquet fuzzy integral fusion," *Inf. Fusion*, vol. 3, no. 1, pp. 69–85, 2002.
- [10] P. D. Gader, B. Nelson, A. Hocaoglu, S. Auephanwirayakul, and M. Khabou, H. Bunke and A. Kandel, Eds., "Neural versus heuristic development of Choquet fuzzy integral fusion algorithms for land mine detection," in *Neuro-Fuzzy Pattern Recognition*. Singapore: World Scientific, 2000, pp. 205–226.
- [11] M. Grabisch, "Fuzzy integral for classification and feature extraction," in *Fuzzy Measures and Integrals. Theory and Applications*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds. Heidelberg, Germany: Springer-Verlag, 2000, pp. 348–374.
- [12] J. M. Keller, P. D. Gader, H. Tahani, J. H. Chiang, and M. Mohamed, "Advances in fuzzy integration for pattern recognition," *Fuzzy Sets Syst.*, vol. 65, no. 1, pp. 273–283, 1994.
- [13] H. Tahani and J. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans. Syst. Man, Cybern.*, vol. 20, no. 3, pp. 733–741, May/Jun. 1990.
- [14] M. Grabisch and M. Sugeno, "Multi-attribute classification using fuzzy integral," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Mar. 1992, pp. 47–54.
- [15] J.-H. Chiang, "Choquet fuzzy integral-based hierarchical network for decision analysis," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 1, pp. 63–71, Feb. 1999.
- [16] A. Hocaoglu, P. Gader, and J.-H. Chiang, "Comments on Choquet fuzzy integral-based hierarchical network for decision analysis," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 48–53, Dec. 1999.
- [17] M. Grabisch, T. Murofushi, and M. Sugeno, *Fuzzy Measures and Integrals. Theory and Applications*, ser. Studies in Fuzziness and Soft Computing. Heidelberg, Germany: Springer-Verlag, 2000.
- [18] P. Gader, L. Wen-Hsiung, and A. Mendez-Vazquez, "Continuous Choquet integrals with respect to random sets with applications to landmine detection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2004, vol. 1, pp. 523–528.
- [19] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Dept. Physics, Tokyo Inst. Technol., Tokyo, Japan, 1974.
- [20] Z. Wang and G. J. Klir, *Fuzzy Measure Theory*. Norwell, MA: Kluwer, 1993.
- [21] K. Leszczynski, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measure and fuzzy clustering," *Fuzzy Sets Syst.*, vol. 15, no. 2, pp. 147–158, Mar. 1985.
- [22] M. Grabisch and J. Nicolas, "Classification by fuzzy integral: Performance and tests," *Fuzzy Sets Syst.*, vol. 65, no. 2-3, pp. 255–271, 1994.
- [23] B.-H. Juang, W. Chou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 257–266, May 1997.
- [24] S. Katagiri, B.-H. Juang, and C.-H. Lee, "Pattern recognition using a family of design algorithms based upon the generalized probabilistic descent method," *Proc. IEEE*, vol. 86, no. 11, pp. 2345–2372, Nov. 1998.
- [25] M. G. Rahim, B.-H. Juang, and C.-H. Lee, "Discriminative utterance verification for connected digit recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 266–277, May 1997.
- [26] C. Wen-Tsong and P. Gader, "Word level discriminative training for handwritten word recognition," in *Proc. 7th Int. Workshop Frontiers Handwritten Recognit.*, Sep. 2000, pp. 393–402.
- [27] H. Mizutani, "Discriminative learning for error and minimum reject classification," in *Proc. 14th Int. Conf. Pattern Recognit.*, 1998, vol. 1, pp. 136–140.
- [28] P. D. Gader, R. Grandhi, W.-H. Lee, and J. N. Wilson, "Integration of ordered weighted averaging operators with feed-forward neural networks for feature subset selection and pattern classification," *IEEE Trans. Fuzzy Syst.*, submitted for publication.
- [29] P. D. Gader, W.-H. Lee, and J. N. Wilson, "Detecting landmines with ground penetrating radar using feature-based rules order statistics, and adaptive whitening," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.
- [30] K. C. Ho, L. Carin, P. D. Gader, and J. N. Wilson, "On using the spectral features from ground penetrating radar for landmine-clutter discrimination," *IEEE Trans. Geosci. Remote Sens.*, submitted for publication.
- [31] H. Frigui and P. D. Gader, "Detection and discrimination of land mines based on edge histogram descriptors and fuzzy k-nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Vancouver, BC, Canada, Jul. 2006, pp. 1494–1499.
- [32] H. Frigui, P. Gader, and K. Satyanarayana, "Detection and discrimination of landmines in ground penetrating radar using an eigenmine and fuzzy membership function approach," in *Proc. SPIE Conf. Detection Remediation Technol. Mines Minelike Targets IX*, Orlando, FL, Apr. 2004, pp. 772–780.
- [33] P. A. Torriane and L. M. Collins, "Application of texture feature classification methods to landmine/clutter discrimination in off-road GPR data," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2004, pp. 1621–1624.
- [34] P. A. Torriane, C. S. Throckmorton, and L. M. Collins, "Performance of an adaptive feature-based processor for a wideband ground penetrating radar system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 2, pp. 644–659, Apr. 2006.
- [35] T. H. Gormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*. New York: McGraw-Hill, 2001.
- [36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge Univ. Press, 1992.
- [37] G. Klir and Z. Wang, *Fuzzy Measure Theory*. New York: Plenum, 1992.



Andres Mendez-Vazquez is currently working towards the Ph.D. degree in computer engineering at the University of Florida, Gainesville.

His research interests include landmine detection, statistical methods for machine learning, and fuzzy measures and Choquet integration.



Paul Gader (SM'99) received the Ph.D. degree in mathematics from the University of Florida, Gainesville, in 1986.

He has worked as a Senior Research Scientist at Honeywell's Systems and Research Center, as a Research Engineer and Manager at the Environmental Research Institute of Michigan, and as a faculty member at the Universities of Wisconsin, Missouri, and Florida, where he is currently a Professor of Computer and Information Science and Engineering.

He led teams involved in real-time, handwritten address recognition systems for the U.S. Postal Service and teams that devised and tested several real-time algorithms in the field for mine detection. He has over 165 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including over 55 refereed journal articles. His research interests include landmine detection, handwriting recognition, mathematical morphology, machine learning, and fuzzy sets and Choquet integration.



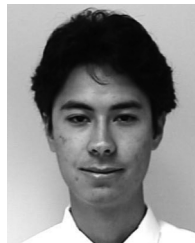
James M. Keller (F'00) received the Ph.D. degree in mathematics from the University of Missouri-Columbia, Columbia, in 1978.

He holds the University of Missouri Curators' Professorship in the Electrical and Computer Engineering and Computer Science Departments on the Columbia campus. He is also the R. L. Tatum Professor in the College of Engineering. His research interests center on computational intelligence: fuzzy set theory and fuzzy logic, neural networks, and evolutionary computation, with a focus on problems

in computer vision, pattern recognition, and information fusion, including bioinformatics, spatial reasoning in robotics, sensor and information analysis in technology for eldercare, and landmine detection. He has been supported by several industrial and government institutions, including the Electronics and Space Corporation, Union Electric, Geo-Centers, National Science Foundation (NSF), the Administration on Aging, NASA/JSC, the Air Force Office of Scientific Research, the Army Research Office, the Office of Naval Research, and the Army Night Vision and Electronic Sensors Directorate. He has coauthored over 250 technical publications.

Dr. Keller is an Associate Editor of the *International Journal of Approximate Reasoning*, and is on the editorial board of *Pattern Analysis and Applications*, *Fuzzy Sets and Systems*, *International Journal of Fuzzy Systems*, and the *Journal of Intelligent and Fuzzy Systems*. He has presented live and video tutorials on fuzzy logic in computer vision with the IEEE, is a National Lecturer for the Association for Computing Machinery (ACM), is an IEEE Computational Intelligence Society Distinguished Lecturer, and was the past President of the North American Fuzzy Information Processing Society (NAFIPS). He finished a full six-year term as an Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. Currently, he is the Vice President for Publications of the IEEE Computational Intelligence Society. He was the Conference Chair of the

1991 NAFIPS Workshop, Program Cochair of the 1996 NAFIPS meeting, Program Cochair of the 1997 IEEE International Conference on Neural Networks, and the Program Chair of the 1998 IEEE International Conference on Fuzzy Systems. He was the General Chair for the 2003 IEEE International Conference on Fuzzy Systems.



Kenneth Chamberlin

He received the B.S. and M.S. degrees in computer engineering from the University of Florida, Gainesville, in 2004.

Currently, he is with CAE USA, Tampa, FL, where he specializes in the development and integration of visual systems for military flight simulators.

A Large-Scale Systematic Evaluation of Algorithms Using Ground-Penetrating Radar for Landmine Detection and Discrimination

Joseph N. Wilson, *Member, IEEE*, Paul Gader, *Senior Member, IEEE*, Wen-Hsiung Lee, Hichem Frigui, and K. C. Ho, *Senior Member, IEEE*

Abstract—A variety of algorithms for the detection of landmines and discrimination between landmines and clutter objects have been presented. We discuss four quite different approaches in using data collected by a vehicle-mounted ground-penetrating radar sensor to detect landmines and distinguish them from clutter objects. One uses edge features in a hidden Markov model; the second uses geometric features in a feed-forward order-weighted average network; the third employs spectral features as its basis; and the fourth clusters edge histograms. We present the results of a large-scale cross-validation evaluation that uses a diverse set of data collected over 41 807.57 m² of ground, including 1593 mine encounters. Finally, we discuss the results of that ranking and what one can conclude concerning the performance of these four algorithms in various settings.

Index Terms—Discrimination, ground-penetrating radar (GPR), landmine detection.

I. INTRODUCTION

GROUND-PENETRATING radar (GPR) sensors have been used in a variety of landmine detection systems for quite some time [1], and various computer algorithms in processing GPR data to detect landmines and discriminate between landmines and nonmine clutter objects have been employed [2]–[15]. Systematic evaluations and comparisons of these algorithms are rare, however. Our purpose here is to present the results of an evaluation of four different landmine discrimination algorithms that are applied to data collected with a vehicle-mounted radar system over 41 807.57 m² of ground.

Manuscript received September 1, 2006; revised May 10, 2007. This work was supported in part by the Army Research Office under Grant W911NF-05-1-0067. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

J. N. Wilson and P. Gader are with the Computer and Information Science and Engineering Department, University of Florida, Gainesville, FL 32611 USA (e-mail: jnw@cise.ufl.edu; pgader@cise.ufl.edu).

W.-H. Lee is with NIITEK, Inc., Sterling, VA 20166 USA (e-mail: wlee@NIITEK.com).

H. Frigui is with the Multimedia Research Laboratory, University of Louisville, Louisville, KY 40202 USA (e-mail: h.frigui@louisville.edu).

K. C. Ho is with the Electrical and Computer Engineering Department, University of Missouri, Columbia, MO 65211 USA (e-mail: hod@missouri.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2007.900993

A NIITEK, Inc., landmine detection system comprising a vehicle-mounted 24-channel GPR array [16], [17] was used to collect data from a variety of test sites. The sites include dirt and gravel roads and lanes and contain both landmines and clutter objects. The data collected by the NIITEK system are used as input to each of the detection algorithms. The NIITEK GPR collects 24 channels of data. Adjacent channels are spaced approximately 5 cm apart in the crosstrack direction. The downtrack interval between samples in each channel is approximately 5 cm. The system uses a V-dipole antenna that generates a wideband pulse ranging from 200 MHz to 7 GHz. Each A-scan, that is, the measured waveform that is collected in one channel at one downtrack position, contains 416 time samples at which the GPR signal return is recorded. Each sample corresponds to roughly 8 ps. Although we often refer to the time index as depth, since the radar wave is traveling through different media, this index does not represent a uniform sampling of depth. Thus, we model an entire collection of input data as a 3-D matrix of sample values $S(x, y, z)$, where the indices x , y , and z represent downtrack position, crosstrack position, and depth, respectively.

Fig. 1 shows several B-scans (sequences of A-scans) of both downtrack (formed from a time sequence of A-scans from a single sensor channel) and crosstrack (formed from each channel's response in a single sample). The surveyed object position is highlighted in each figure. The objects scanned are the following: 1) a high-metal content antitank mine; 2) a low-metal antitank mine; 3) a soft-drink can; and 4) a wood block.

II. DISCRIMINATION ALGORITHMS

Landmine detection algorithms, like many other target detection algorithms, typically consist of a number of discrete phases. Often, a prescanner is applied to reduce the volume of data to be inspected by later phases. The prescanner identifies distinct alarms (points of interest) in the data. Features are then extracted from the data corresponding to the alarms. Then, these features are presented to an algorithm that discriminates between landmines and nonmine objects (false alarms). We are concerned here in evaluating the utility of discrimination algorithms.

Various algorithms have been applied to the problem of discrimination between landmines and false alarms. In this paper, we consider four specific algorithms of distinct character. The

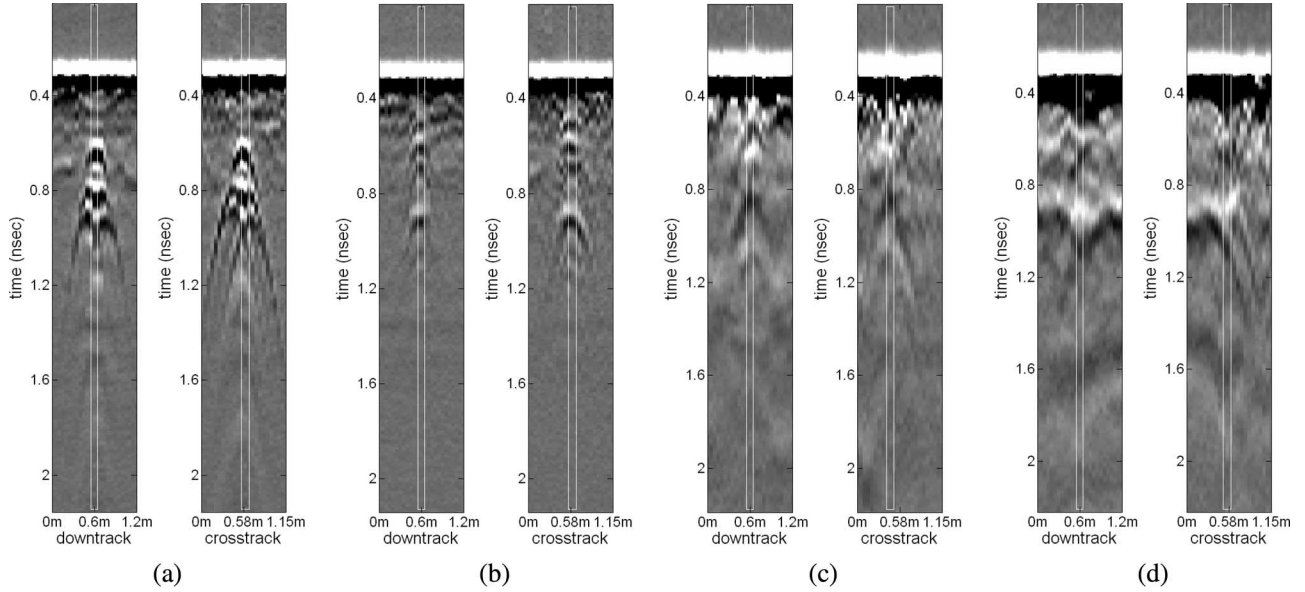


Fig. 1. NIITEK Radar downtrack and crosstrack B-scans pairs. (a) Metal mine. (b) Low-metal mine. (c) Soft-drink can. (d) Wood block.

first employs a hidden Markov model (HMM) that models the time-varying behavior of GPR signals encoded using edge direction information to compute the likelihood that a sequence of measurements is consistent with a buried landmine. The second extracts geometric features of the GPR data associated with a ground location and applies a feed-forward order-weighted average (FOWA) network to discriminate between landmines and clutter. The third algorithm extracts features from the frequency spectrum of the GPR data associated with a ground location and formulates a confidence value based on similarity to a collection of features that characterize mine objects. The final algorithm extracts edge histograms capturing the frequency of occurrence of edge orientations in the data associated with a ground position and then uses a fuzzy K-nearest neighbor (K-NN) algorithm to generate a mine confidence level.

Rather than processing each of the many discrete locations sampled by the GPR array, our algorithms restrict their processing to alarm locations identified by a prescreener algorithm. The prescreener can be thought of as a conservative detection algorithm, that is, one designed to provide a high probability of landmine detection at the expense of inclusion of many false alarms. False alarms arise as a result of radar signals that present a minelike character. Such signals are generally said to be a result of clutter. In this evaluation, clutter arises from two different processes. One type of clutter is emplaced and surveyed in an effort to test the robustness of the algorithms. Other clutter is a result of either human activity unrelated to the data collection or natural processes. We refer to this second kind of clutter as nonemplaced. Nonemplaced clutter includes objects discarded or lost by humans, soil inconsistencies and voids (due to formation processes, erosion, or excavation), stones, roots, and other vegetation, as well as remnants of animal activity. It is the job of the subsequent algorithms to discriminate between those prescreener alarms corresponding to landmines and those corresponding to clutter. All algorithms considered here were applied to data that were prescreened

using the Duke University NUKEv6 prescreener, a variant of the least mean square prescreener [18], [19]. A version of this algorithm (F1) has been implemented in real time in a uniprocessor system. The prescreener detected 1560 of 1593 mines encountered in the data, yielding a 97.9% probability of detection. It rejected 161 of 211 emplaced clutter objects encountered. It yielded a total of 3435 false alarms that are associated with nonemplaced clutter objects.

A. HMM Algorithm

The NIITEK GPR system produces sequences of observation vectors that can be considered as functions of uniform time (and space if the vehicle velocity is constant). Signals arising from the presence of buried landmines can be used to develop an HMM that captures the probabilities that sequences of these signals were produced by landmines and to infer the location of possible landmines. We modified the work of Frigui *et al.* [25] to give us an HMM suitable for use with the NIITEK GPR data.

HMMs are stochastic models for complex processes that produce time sequences of random observations as a function of states. They have been successfully applied to the problems of speech and handwriting recognition [20]–[22]. An HMM produces a sequence of random observation vectors at discrete times according to an underlying Markov chain. At each observation time, the Markov chain may be in one of N states, and given that the chain is in a certain state, there are probabilities of moving to other states. These probabilities are called the transition probabilities.

The model is said to be hidden because the states are not directly observable. Given an observation vector at time t , and a state S , there is a probability that the chain is in state S . The actual state is described by a probability density function, which can either be continuous or discrete. The probability density functions describing the states define the probabilities of the

observations conditioned upon the chain being in the associated state. Thus, the HMM is characterized by three sets of probability density functions: the transition probabilities, the state probability density functions, and the initial probabilities. In the case of the discrete HMM, the observation vectors are typically quantized into a finite set of symbols, called the codebook. Each state is represented by a discrete probability density function that assigns each symbol a probability of occurring given that the system is in a given state.

We use Rabiner's notation [21], [22] in the brief discussion here. The compact notation $\lambda = (A, B, \pi)$ is used to indicate the parameter set of an HMM, where $A = \{a_{i,j}\}$, $a_{i,j} = P(q_{t+1} = S_j | q_t = S_i)$ are the state transition probability distributions, $B = \{b_j(k)\}$, $b_j(k) = P(\nu_k \text{ at } t | q_t = S_j)$ are the observation symbol probabilities (of encountering observation k in state j), and $\pi = \{\pi_i\}$, $\pi_i = P(q_1 = S_i)$ are the initial state probabilities.

The three problems of interest that must be solved to employ the model are as follows: 1) classification; 2) identifying an optimal state sequence; and 3) estimating the model parameters.

Classification involves computing the probability of an observation sequence $O = O_1, O_2, \dots, O_T$ given a model λ , $P(O|\lambda)$. In the landmine detection problem, this corresponds to finding the probability of observing a sequence of GPR signals when the sequence is associated with a mine and λ is a landmine model; or when the sequence is a result of clutter and λ is a clutter model.

In applications, it often turns out that computing an optimal state sequence is more useful than $P(O|\lambda)$. There are several possible ways of finding an optimal state sequence associated with the given observation sequence, depending on the definition of the optimal state sequence, i.e., there are several possible optimality criteria. One that is particularly useful is to maximize $P(O, Q|\lambda)$ over all possible state sequences Q . The Viterbi algorithm is an efficient formal technique in finding this maximum state sequence and associated probability.

The Baum–Welch algorithm [23], [24], which is an iterative approach to parameter estimation, was used to identify the parameters of the model employed in this paper. The parameters for the model employed in this paper were created using a different radar system [25].

Our goal is to produce a scalar value indicating our confidence that a buried landmine is present at any of the various spatial positions (x, y) encountered by the vehicle-mounted sensor. To fit into the HMM context, a sequence of observation vectors must be produced for each point. These observation vectors are features that encode important information about the landmine signatures in a compact form. The downtrack observation sequence at the point (x, y) will be the sequence of observation vectors $O(x, y - k), O(x, y - k + 1), \dots, O(x, y - 1), O(x, y), O(x, y + 1), \dots, O(x, y + k)$, and the crosstrack sequence is the set of vectors $O(x - k, y), O(x - k + 1, y), \dots, O(x - 1, y), O(x, y), O(x + 1, y), \dots, O(x + k, y)$. To generate these observations, we preprocess the data to accentuate edges in the diagonal and antidiagonal directions. Let $S(x, y, z)$ denote the raw 3-D GPR data. The downtrack and crosstrack second derivatives are first estimated on the raw data. The reason for differentiating is that it removes stationary

effects that remain relatively constant from scan to scan such as the return from the ground and the standing pattern caused by the interaction of the GPR with the surrounding components. Although differentiation is sensitive to noise, the NIITEK data are not very noisy; thus, clutter objects rather than system noise will be more likely to yield false alarms. The features calculated from this second-derivative images are the strengths of diagonal and antidiagonal edges calculated from downtrack or crosstrack B-scans.

The discrete mine model has three states as does the background model. The discrete mine model is a left-to-right model, in that, states are ordered, and the transition probabilities in moving to a lower numbered state are zero. The three mine states correspond to the leading edge, center, and trailing edge of a mine. Two optimal state sequences are computed for the mine model. One assuming the model is in the third mine state at the final time, and the other assuming the model is in the background state at the final time. The state sequence with the highest probability produces the model output x (the downtrack response) and y (the crosstrack response). These are combined to form the HMM score $h = (\alpha x + (1 - \alpha)y) + \sqrt{xy}$, where α is chosen to be 0.5 for alarms in channels 6–19, and 0.75 for channels 1–5 and 20–24. This assigns equal weight to the individual crosstrack and downtrack responses in those channels in which most of the mine signature is expected to be fully present in the crosstrack scans, and a higher weight to the downtrack response in those channels near the edges of the data volume where only a portion of the crosstrack sequence is expected to appear. Finally, the geometric mean of the combined downtrack, crosstrack HMM response, and the prescreener confidence p is used as the resulting mine confidence $\text{Conf} = \sqrt{hp}$.

We can summarize the HMM algorithm processing steps as follows.

- 1) Estimate downtrack and crosstrack second derivative B-scans.
- 2) Form observation sequences from diagonal and antidiagonal edge features in second derivative B-scans.
- 3) Find mine model probabilities x and y using downtrack and crosstrack observation sequences, respectively.
- 4) Form HMM score $h = (\alpha x + (1 - \alpha)y) + \sqrt{xy}$ and confidence $\text{Conf} = \sqrt{hp}$.

B. Geometric Feature FOWA ROCA Algorithm (GEOM)

The GEOM is based on a single hidden-layer FOWA network [30], which is essentially a perceptron with a combination of scalar and order-weighted average vector input features. The features presented to this network are the geometric features of the FROSAW landmine detection algorithm [27]. To improve the algorithm's accuracy, we employ an iterative technique that maximizes the area under the receiver operating characteristics (ROC) curve, which we refer to as ROCA [28].

The features employed by this algorithm are geometric features of the GPR data. These features are captured in a depth-bin whitened version of the GPR data. The GPR data are segmented into a sequence of subimages that overlap in the depth dimension. To reduce noise, decorrelate time samples,

and reduce computational burden, principal component analysis is used to reduce the number of elements in depth bins on a channel-by-channel basis.

It has been consistently observed that in many of the depth bins, the whitened energy signal for mines has a compact, solid, and circular shape (sometimes also accompanied by outer rings). On the other hand, whitened energy signals for nonmine-like false alarms (i.e., those alarms having raw GPR signatures that humans qualitatively label as nonminelike) tend to be irregular. Based on these observations, the following features are computed from the whitened energy signals for discriminating mines and nonmines: compactness, eccentricity, solidity, area/filled area ratio.

To gauge the compactness of a whitened energy signal, two approaches from the FROSAW algorithm [27] are taken. Both approaches measure the compactness centered at an alarm location. The first approach is referred to as adaptive compactness, whereas the second approach is referred to as fixed compactness. Adaptive compactness is defined as the radius from the centroid required for a region of that radius to contain a fixed percentage of the energy of a relatively large radius region. More precisely, let (x_a, y_a) denote the location of the alarm under consideration and let $e_w(x, y, z)$ be the whitened energy associated with the alarm. The whitened energy is normalized as follows:

$$\tilde{e}_w(x, y, z) = \frac{e_w(x, y, z) - \mu_s}{\sigma_s} \quad (1)$$

where μ_s and σ_s are the mean and standard deviation over all whitened energy values associated with the alarm. Denote the normalized whitened energy of the n th depth bin within the disk of radius r by

$$E_n(r; x_a, y_a) = \sum_{(x,y) \in D(r; x_a, y_a)} \tilde{e}_w(x, y, z)^2 \quad (2)$$

where $D(r; x_a, y_a)$ is the disk of radius r centered at alarm location (x_a, y_a) . Let $r_{\max} > 1$ denote a fixed radius and E_p an energy threshold. The adaptive compactness at depth n_0 at location (x_a, y_a) is defined as

$$p_{n_0}(x_a, y_a) = 1/\min \left\{ r : 1 \leq r \leq r_{\max} \text{ and } \frac{E_{n_0}(r; x_a, y_a)}{E_{n_0}(r_{\max}; x_a, y_a)} \geq E_p \right\}. \quad (3)$$

Fixed compactness is defined as the ratio of the energy in a 5×5 region to the energy in a 24×25 region, both regions being centered at the reported alarm location in the downtrack direction. That is, the fixed compactness at depth n_0 at location (x_a, y_a) is defined as

$$p_{n_0 f}(x_a, y_a) = \frac{E_{\text{inner}}(x_a, y_a)}{E_{\text{outer}}(x_a, y_a)} \quad (4)$$

where

$$E_{\text{inner}}(x_a, y_a) = \sum_{x=x_a-2}^{x_a+2} \sum_{y=y_a-2}^{y_a+2} \tilde{e}_w(x, y, z_0)^2 \quad (5)$$

$$E_{\text{outer}}(x_a, y_a) = \sum_{x=1}^{24} \sum_{y=y_a-12}^{y_a+12} \tilde{e}_w(x, y, z_0)^2. \quad (6)$$

In general, mines have larger values of compactness than false alarms not associated with emplaced clutter.

To compute additional features of the normalized whitened energy signal $\tilde{e}_w(x, y, z)$, the signal is first thresholded using Otsu's method [29]. After thresholding, connected components are formed. Only the connected component with gray-level centroid closest to the reported alarm location is kept for computing features on the z_0^{th} depth bin. The additional features of this component region (eccentricity, solidity, and ratio of area to filled area) are computed as in the FROSAW algorithm [27].

The FOWA algorithm employs vectors of these depth features by computing an order-weighted average (OWA) of them. An OWA operator [31]–[33] $F: R^n \rightarrow R^n$ has a weight vector $W = [w_1, \dots, w_j]$ satisfying $\sum_{j=1}^n w_j = 1$ and such that $F(a_1, \dots, a_n) = \sum_{j=1}^n w_j a_{(j)}$, where $a_{(j)}$ is the j th largest of the a_i . The input to the FOWA network comprises I feature values, I_0 of which are vector-valued and the rest having scalar values. Each element of this geometric GPR FOWA network is a feature calculated on a single depth bin; therefore, for example, the solidity feature contains an entry for each whitened energy depth-bin's Otsu-thresholded region solidity. Thus, inputs, $\alpha_m = [\alpha_{m,1}, \alpha_{m,2}, \dots, \alpha_{m,K_m}]^T$, $m = 1, 2, \dots, I_0$, are vector-valued features, and $\alpha_{I_0+1}, \alpha_{I_0+2}, \dots, \alpha_I$ have scalar values, and the whole collection of features is $\mathbf{z} = [\alpha_1^T, \alpha_2^T, \dots, \alpha_{I_0}^T, \alpha_{I_0+1}, \alpha_{I_0+2}, \dots, \alpha_I]^T$. First, I_0 OWA operators are applied, one to each of the vector-valued features. The output of this layer is a vector $\lambda_m = \sum_{k=1}^{K_m} w_{m,k} \alpha_{m(k)}$, $m = 1, 2, \dots, I_0$. For $m = I_0 + 1, I_0 + 2, \dots, I$, $\lambda_m = \alpha_m$, that is, these features are not sorted and weighted by the OWA operators. With tanh sigmoid functions being employed at the hidden and output layers, the outputs at the hidden layer and output layer are, respectively

$$h_l = \tanh \left(\beta_1 \sum_{m=1}^I w_{l,m}^h \lambda_m \right) \quad (7)$$

$$f(\mathbf{z}; \theta) = \tanh \left(\beta_2 \sum_{l=1}^L w_l^z h_l \right) \quad (8)$$

where L is the number of hidden nodes, and θ is a vector with all the weights $\{w_{m,k}\}$, $\{w_{l,m}^h\}$, and $\{w_l^z\}$ as its elements. In our notation, \mathbf{z} can be either \mathbf{x}^i for mines or \mathbf{y}^j for nonmines.

We initially train the FOWA network by minimizing the mse between $f(\mathbf{x}^i; \theta)$ and the desired output for mine objects, and the mse between $f(\mathbf{y}^j; \theta)$ and the desired output for nonmine objects, namely, false alarms not associated with emplaced clutter. However, after performing this training, a second iterative technique optimizes an objective function that seeks to maximize the area under the ROC curve using a steepest

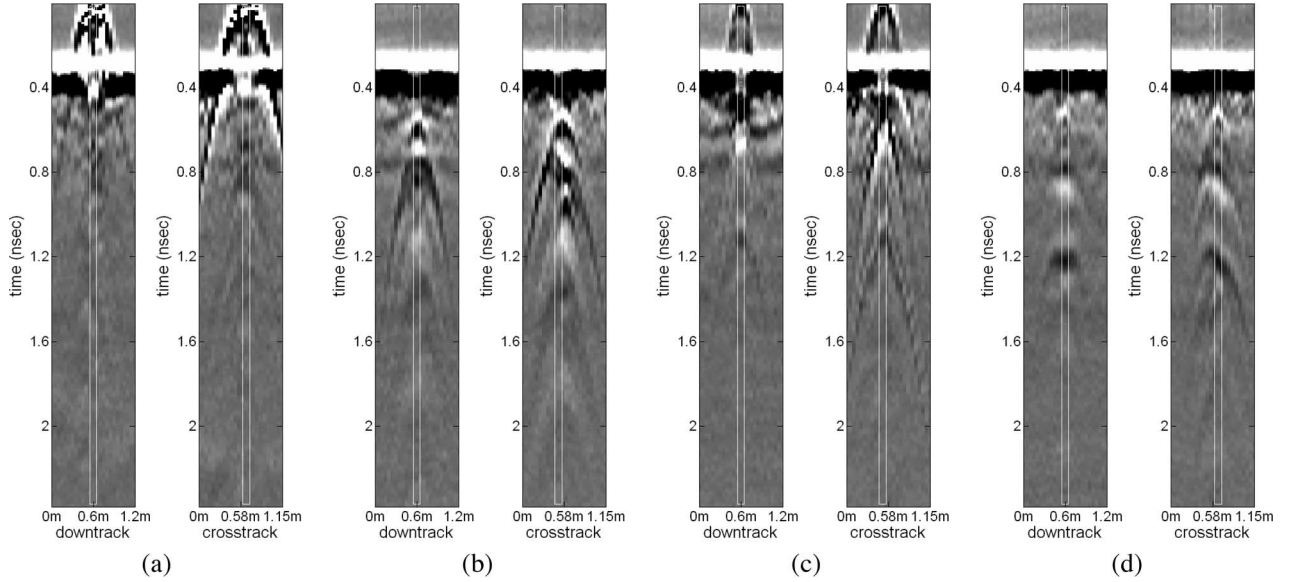


Fig. 2. Flush-buried and surface-laid mine signatures. (a) Surface metal mine. (b) Flush metal mine. (c) Surface low-metal mine. (d) Flush low-metal mine.

descent method [28]. Briefly, this technique attempts to adjust the parameters of the objective function by considering those mine alarms and false alarms whose confidences are within a small distance Δt of each other. Let C denote the collection of indices (i, j) of pairs of mines \mathbf{x}^i and false alarms \mathbf{y}^j falling into this category. Then, our steepest descent method adjusts the objective function parameter w_l by $dw_l = (s/\Delta t) \sum_{(i,j) \in C} \times ((\partial f(\mathbf{x}^i; \theta)/\partial w_l) - (\partial f(\mathbf{y}^j; \theta)/\partial w_l))$, where s is a heuristically determined step size. That is, it uses the summed weighted differences of the confidences of similarly scored mines and false alarms to increase their difference. The confidence reported is the output of the network evaluated with adjusted parameters θ , $\text{Conf} = f(\mathbf{z}; \theta)$.

In summary, the geometric FOWA ROCA algorithm processing steps are as follows.

- 1) Generate whitened depth-bin volumes.
- 2) Compute geometric features from each depth bin, $g_i(j)$ being a feature value i at depth j .
- 3) Apply the FOWA ROCA network f to the geometric features $\mathbf{z} = (\mathbf{g}_1, \dots, \mathbf{g}_n)$ using training set parameters θ to yield $\text{Conf} = f(\mathbf{z}; \theta)$.

C. Spectral Confidence Feature Algorithm

In contrast to the geometric features and the edge histogram features, the spectral confidence feature algorithm (SCF) aims at capturing characteristics of a target in the frequency domain. The spectral feature is derived from the energy density spectrum (EDS) of an alarm declared by the prescreener. The estimation of EDS involves four steps: 1) preprocessing; 2) nonlinear smoothing; 3) whitening; and 4) averaging.

Preprocessing estimates the ground level, aligns the data from each scan with respect to the ground level, and applies range gating to remove the data above and near the ground surface. Subpixel alignment with a step of 0.25 pixels is applied to obtain better alignment, and the range gating removes the data from the start until 20 depth pixels below the ground

level. Range gating is necessary; otherwise, the EDS will be dominated by the response resulted from the ground bounce. Fig. 2. shows b-scans of both flush-buried and surface-laid metal and plastic mines. The presence of signal associated with pixels more than 20 samples (0.16 ns) below the initial ground bounce provides the opportunity to identify these mines after range gating.

The whitening step performs equalization on the spectrum from the background so that the estimated EDS reflects the actual spectral characteristics of an alarm. Let $D(x, y, k_z)$ be the Fourier transform of the data along depth at the position (x, y) , where k_z denotes the frequency index. The mean and the standard deviation of the background are estimated at each crosstrack and each frequency index from the past downtrack samples as

$$m(x, k_z) = \frac{1}{L} \left(\sum_{i=y_o-G-L}^{y_o-G-1} D(x, i, k_z) \right)$$

$$\sigma(x, k_z) = \sqrt{\frac{1}{L} \left(\sum_{i=y_o-G-L}^{y_o-G-1} |D(x, i, k_z)|^2 \right) - |m(x, k_z)|^2}$$
(9)

where (x_o, y_o) is the alarm location declared by the prescreener, $G = 6$ is the number of guard samples that avoid the use of target samples, and $L = 58$ is the number of samples that estimate the background statistics. The spectral whitening is achieved by the normalization

$$\tilde{D}(x, y, k_z) = \left(\frac{D(x, y, k_z) - m(x, k_z)}{\sigma(x, k_z)} \right),$$

$$y = y_o - G, y_o - G + 1, \dots, y_o + G. \quad (10)$$

Note that $\tilde{D}(x, y, k_z)$ is complex. The mean and root-mean-square (rms) value of the magnitude $|\tilde{D}(x, y, k_z)|$ are next

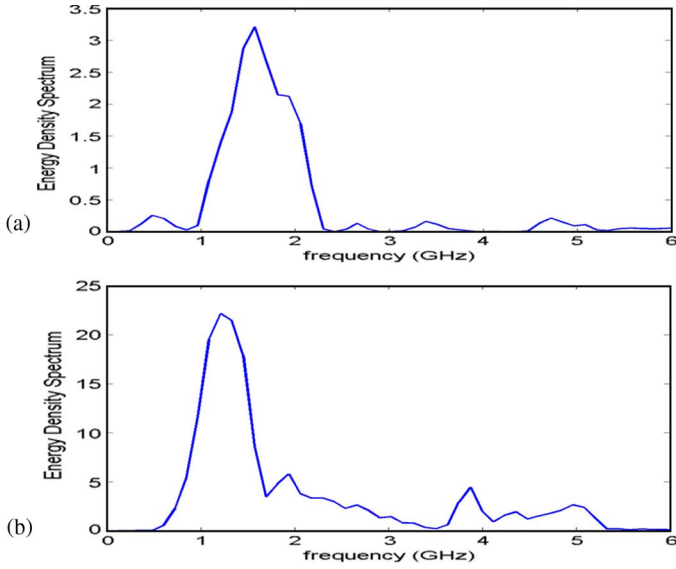


Fig. 3. EDS of two low-metal differing antitank mines.

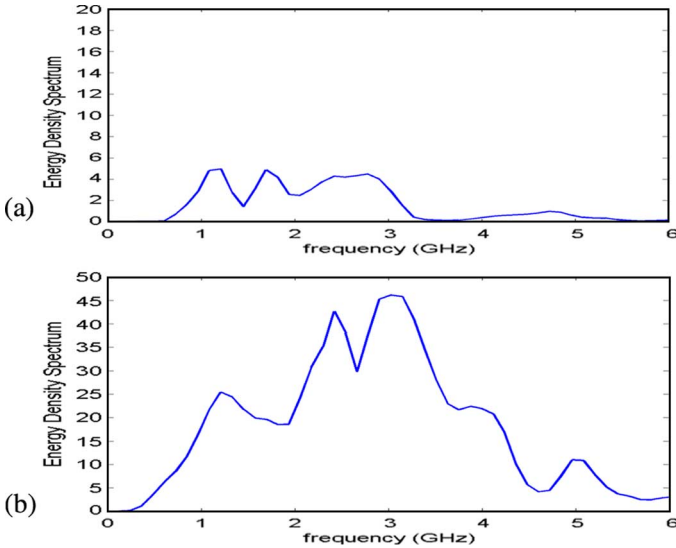


Fig. 4. EDS of two different types of clutter targets (a) metal debris and (b) plastic clutter.

computed over $y = y_o - G, y_o - G + 1, \dots, y_o + G$. Size-contrast processing by subtracting the mean and setting to zero the values less than the rms value is applied, resulting in $U(x, y, k_z)$.

Averaging reduces the variance in the EDS by forming

$$P(x_o, y_o, k_z) = \frac{1}{N^2} \sum_{x=x_o-(N-1)/2}^{x_o+(N-1)/2} \sum_{y=y_o-(N-1)/2}^{y_o+(N-1)/2} U(x, y, k_z) \quad (11)$$

where the averaging is over 25 cm downtrack and 25 cm crosstrack, which corresponds to $N = 5$. $P(x_o, y_o, k_z)$ is the EDS estimate to be used in extracting the spectral features.

Fig. 3 depicts the EDS of two low-metal antitank mine targets of different types, and Fig. 4 shows the EDS of a metal object (a 5×5 cm spool of resin-core solder) and a plastic clutter (a 12-cm diameter container lid together with an 8-cm container

lid). The EDS produced by these mines and clutter objects are obviously different, motivating the use of this feature for discrimination. We must point out that, although this difference we have observed in EDS between mines and clutter is present for a wide variety of objects, there are many clutter object signals whose EDS is quite similar to that of the mines shown. Likewise, there are mine signals whose EDS do not so closely resemble those shown in Fig. 3.

The spectral peaks from mine targets could vary between 1.2 and 2 GHz. Subbanding using a cosine square window with 50% overlap is applied, where each subband is 600 MHz. The spectral energy in each subband is computed by summing the EDS values within the subband, resulting in ten values, denoted by a column vector Q , over a 6-GHz range. Based on the matched filtering approach, we then calculate the dot product between Q and seven spectral masks that are derived through training from mine targets. Let W be the spectral mask that gives the largest dot product. The spectral feature value used in this paper is $\text{Conf} = (\log(\mathbf{W}^T \mathbf{Q} + 1) + k)(p - p_{\min})$. This confidence value geometrically combines the spectral confidence with prescreener confidence p . The log operation reduces the dynamic range of the spectral confidence value, $k = 1.5$ is used to let the prescreener confidence dominate when the spectral confidence is low, and p_{\min} , the prescreener threshold value, is subtracted to make the prescreener confidence value be zero-based.

We can summarize the spectral feature algorithm as follows.

- 1) Perform ground alignment and range gating.
- 2) Set $D(x, y, k_z)$ to the Fourier transform along depth of the data volume.
- 3) Whiten D based on background samples.
- 4) Perform size contrast processing on D yielding $U(x, y, k_z)$.
- 5) Find the mean depth vector value of U in a 25×25 cm neighborhood.
- 6) Sum U within ten frequency bands to form Q .
- 7) Find $\text{Conf} = (\log(\mathbf{W}^T \mathbf{Q} + 1) + k)(p - p_{\min})$ using the best matching spectral vector \mathbf{W} from a set formed during training.

D. Edge Histogram Discrimination Algorithm

The edge histogram discrimination algorithm [34] uses edge histogram descriptor (EHD) features and employs a rule based on fuzzy K-NNs to assign confidence. A set of alarms with known ground truth is used to train the decision-making process. These labeled alarms are clustered to identify a small number of representatives that capture signature variations due to differing soil conditions, mine types, weather conditions, and so forth. Fuzzy memberships are assigned to these prototypes to capture their degree of similarity to mine and clutter class objects.

The MPEG-7 EHD [35] is used as a feature representation for GPR alarm signatures. The EHD is a mature technique to represent the frequency and the direction of intensity changes appearing within an image. Edges detected within an image are grouped by the EHD into five categories: vertical, horizontal, diagonal (45° rising), antidiagonal (45° falling), and isotropic

(unoriented). The EHD contains five histogram bins counting the number of locations at which each of these edge characterizations dominates the others.

To apply EHD to 3-D GPR data, it is modified to compute two distinct types of 2-D edges, namely, those edges in both downtrack and crosstrack B-scans of the radar data. Let $S_{z,y}^{(x)}$ denote the x th plane of the 3-D signature $S(x, y, z)$. For each $S_{z,y}^{(x)}$, we compute four categories of edge strengths: vertical, horizontal, diagonal, and antidiagonal. If the edge strength in a given direction exceeds threshold θ_G , then the corresponding pixel is considered to be an edge pixel in that direction. Otherwise, it is considered to be an isotropic pixel. We consider images of fixed size at each alarm location (x, y) , spanning $S_{z,y}^{(x')}$ for $x' \in \{x - \delta, x + \delta\}$ and divide these subimages into four horizontally overlapping subimages $S_{z,y_i}^{(x')}$ for $i = \{1, \dots, 4\}$. We compute a five-bin edge histogram H_{z,y_i}^x with bins corresponding to the number of occurrences of each of the assignments of edge to the pixels in subimage $S_{z,y_i}^{(x')}$. Finally, we construct the downtrack component of the EHD, EHD^d , which is defined to be the concatenation of the seven five-bin histograms

$$\text{EHD}^d(S_{xyz}) = [\bar{H}_{zy1} \bar{H}_{zy2} \bar{H}_{zy3} \bar{H}_{zy4} \bar{H}_{zy5} \bar{H}_{zy6} \bar{H}_{zy7}] \quad (12)$$

where $\bar{H}_{zyi} = (1/N_C) \sum_{x=1}^{N_C} H_{zyi}^x$.

To compute the crosstrack EHD component EHD^x , we compute four edge strengths on the $S_{zx}^{(y)}$, $y = 1, \dots, N_S$ planes. Since there are typically fewer crosstrack samples than downtrack samples, we do not divide the crosstrack into subimages. Thus, only one global histogram H_{zx}^y is computed. Otherwise, EHD^x is computed similarly to EHD^d

$$\text{EHD}^x = \frac{1}{N_S} \sum_{y=1}^{N_S} H_{zx}^y. \quad (13)$$

Finally, the composite EHD feature vector is computed as a 40-D histogram that concatenates the downtrack and crosstrack EHD components

$$\text{EHD}(S_{xyz}) = [\text{EHD}^x(S_{xyz}) \text{EHD}^d(S_{xyz})]. \quad (14)$$

A set of labeled alarms with known x, y positions is used as training data. Alarm depths are visually estimated, since the actual depth of a mine or the phenomenon yielding a false alarm cannot be determined by an automated prescreener. Each signature S is a volume cube containing 30 depths, 4 scans, and 7 channels centered at $S_{x,y,z}$, where z is the estimated alarm depth. The training data include signatures of mine alarms and signatures of false alarms not associated with emplaced clutter.

One expects signatures of objects within any given class to exhibit significant variation. Clutter signatures, in particular, may arise from a large number of different types of objects. Mine signatures, as well, may have multiple subclasses corresponding to mines of different types and sizes, buried at different depths, appearing in varying soil and weather conditions, and so forth. Two self-organizing feature maps (one for mines and one for clutter) are used to cluster the alarms. We refer to cluster representatives as prototypes and denote the mine

signature prototypes as R_i^M and the clutter signature prototypes as R_i^C .

Each prototype R_i is assigned a fuzzy membership in each of the class of mines $u^M(R_i)$ and the class of clutter $u^C(R_i)$. We use minimum distance and the Fuzzy C-Means membership function [36] to label new alarms. In particular, for each R_i , we find the closest mine prototype R_i^M and the closet clutter prototype R_i^C , and assign a label using

$$u^M(R_i) = \frac{1/\text{dist}(R_i, R_i^M)}{1/\text{dist}(R_i, R_i^M) + 1/\text{dist}(R_i, R_i^C)}. \quad (15)$$

Each prescreener alarm is tested at multiple depths by sliding the $30 \times 4 \times 7$ EHD window along the depth axis with 50% overlap. At most, ten signatures are extracted for each alarm. The EHD is extracted, and a fuzzy K-NN-based rule is used to assign a confidence value. First, given a test signature S_T , we compute its distance to all representative prototypes. We then sort these distances and identify the K nearest neighbors S_T^1, \dots, S_T^K . Letting p represent the prescreener confidence value, the EHD confidence value is computed as follows:

$$\text{Conf}(S_T) = \left(\frac{\sum_{k=1}^K u^M(S_T^k) \times \frac{1}{\text{dist}(S_T, S_T^k)}}{\sum_{k=1}^K 1/\text{dist}(S_T, S_T^k)} \right)^{1/2}. \quad (16)$$

In summary, one can find the EHD confidence as follows.

- 1) Calculate edge strengths within the downtrack and crosstrack B-scans.
- 2) Form edge histogram features in crosstrack and overlapping downtrack subimages.
- 3) Find the K nearest prototype features.
- 4) Calculate confidence $\text{Conf}(S_T)$ from the test signature's features and the K nearest prototype features as described above.

III. DATASET STATISTICS

The dataset contains data collected between November 2002 and July 2006 from four geographically distinct test sites. Sites A, B, and D are temperate climate test facilities with prepared soil and gravel lanes. Site C is an arid climate test facility with prepared soil lanes. The statistics of the data are shown in Table I. Site B has the largest number of collections and the largest number of alarms. The data collected from Sites B and D have emplaced buried clutter. Although the lanes at Sites A, B, and C were prepared in an attempt to eliminate the presence of minelike objects, they still contain nonemplaced clutter objects. Both metal and nonmetal nonemplaced clutter objects that yielded high mine confidence values such as ploughshares, shell casings, and large rocks were excavated from these sites to determine their nature after the data were collected and their locations had been identified. The emplaced clutter objects include steel scraps, bolts, sort-drink cans, concrete blocks, plastic bottles, wood blocks, and rocks. In all, there are 12 collections having 19 distinct mine types. Many of these mine

TABLE I
STATISTICS OF THE DATASET

Site	Site A	Site B	Site C	Site D	Total
# Collections	3	6	2	1	12
# Mine Types	9	15	9	5	19
# Mine Alarms	183	821	62	494	1560
# Emplaced clutter encounters	0	15	0	196	211
# Emplaced clutter alarms post prescreen	0	4	0	46	50
Area(m ²)	14812.83	15630.62	4054.39	7309.73	41807.57

TABLE II
DISTRIBUTION OF MINE TARGETS AT DIFFERENT DEPTHS

Depth	Surface	0cm	2.5cm	5.1cm	7.6cm	10.2cm	12.7cm	15.2cm	Total
ATLM	12	92	90	204	122	134	47	76	777
ATM	6	37	124	68	151	34	119	77	616
SIM	48		20	47	23	29			167
Total	66	129	234	319	296	197	166	153	1560

types are present at several sites. The data include 1560 mine encounters in a sample ground area of 41 807.57/m².

The distribution of mine targets at different depths is shown in Table II. The targets were buried up to 15.2 cm deep. There were nine distinct types of low-metal antitank mines (ATLMs), 56 high-metal antitank mines (ATMs), and 34 simulants, or simulated mines (SIM).

Fig. 5 shows a histogram of the distribution of mine depths. The mines buried at 2.5–15.2 cm occupy 87.5% of the total targets encountered versus 12.5% surface-laid or flush-buried mines.

IV. EVALUATION

Each of the four algorithms (HMM, GEOM, SCF, and EHD) was implemented for use with the Testing/training Unified Framework system. This system supports creation of supervised learning algorithms that perform discrimination between targets and nontargets in data collected at a variety of different regions (mine lanes) in a variety of different sites. The framework employs algorithms implemented in Matlab using a control flow that incorporates a user-programmed prescreener that processes raw data files into alarms with associated Universal Transverse Mercator coordinates and confidence values. The alarms are then processed by extracting signatures. These signatures are passed to a user-specified feature extractor. The features resulting from the feature extractor are presented along with the alarms to a discrimination algorithm, which produces a confidence for each alarm. The system performs *n*-way cross-validation testing using either lane-based cross-validation

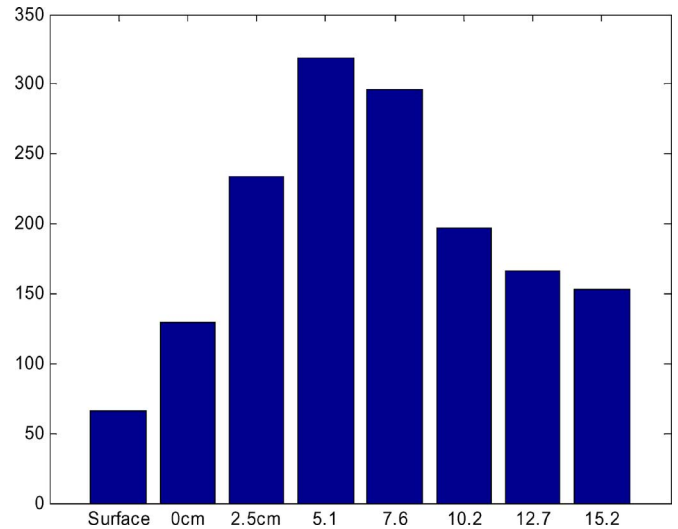


Fig. 5. Distribution of mines at different depths.

(in which each mine lane is, in turn, treated as a test set with the rest of the lanes used for training) or site-based cross-validation (in which each data collection site is treated, in turn, as a test set). The results of this process are scored using the MINE Detection Assessment and Scoring (MIDAS) system developed by Ayers and Rosen of the Institute for Defense Analysis [37]. The GEOM and EHD algorithms are trained in this cross-validation manner. The HMM was based on a model trained using a different radar system [25], and

the SCF employs a single static mine model and is not trained.

Straightforward Matlab implementation of the HMM algorithm requires about five times as much processing time per alarm as does EHD. SCF and FOWA run about eight times as long as the HMM. An efficient C-language implementation of EHD processes a single alarm in 12 ms. Thus, all the algorithms are potentially suited to real-time use.

Various authors have attempted to develop criteria in using ROC curves to compare the performance of algorithms [38], [39]. The work of Ling and Zhang [40] shows that given a constrained environment (in which the number of targets and nontargets is equal) and for a narrowly defined accuracy criterion (best discrimination at median threshold), that maximizing the area under the ROC curve corresponds to increasing accuracy. Provost *et al.* [41], however, argue convincingly that accuracy is not necessarily the best single metric to rank algorithm performance, particularly when comparing ROCs. It is often the case that a single dominating classifier [one producing statistically lower false alarm rate (FAR) at every probability of detection (PD) value] does not exist. Furthermore, in many practical cases such as humanitarian demining, the best algorithm may be the one at which 100% detection is achieved with the lowest false alarm rate, no matter what other properties the ROC may display. For other time-critical demining applications where some level of missed mines is not considered as great a cost, the best ROC may be the one at which the probability of detection is highest at a given constant false alarm rate.

Our algorithm development efforts have been geared toward developing algorithms suitable for an autonomous vehicle-based mine detection system. In any such system, false alarms will delay the progress of the system. To achieve a reasonable rate of progress, we have set an initial goal of reporting fewer than 0.0007 false alarms per square meter at a detection rate of 90%. Our long-term goal is to achieve a false alarm rate below 0.0007/m² at a detection rate of 95%. Knowing, however, that any single property of the ROC may be inappropriate in evaluating the algorithms, we have chosen to consider a number of measurable properties of these ROCs. The metrics chosen for algorithm evaluation are the following:

- 1) PD85: FAR at the threshold yielding PD.85;
- 2) PD90: FAR at PD.90 threshold;
- 3) PD95: FAR at PD.95 threshold;
- 4) FAR0: PD at FAR 0 threshold;
- 5) FAR0.007: PD at FAR 0.0007 threshold;
- 6) FAR0.00007: PAD at FAR 0.00007 threshold;
- 7) SEPAR: Separation of the mine and nonmine confidence distributions, $(\mu_1 - \mu_2)^2 / (\sigma_1^2 + \sigma_2^2)$, where (μ_1, σ_1) are the mean and standard deviation of the mine distribution, and (μ_2, σ_2) are the mean and standard deviation of the nonmine distribution.

Figs. 6–10 show the ROCs associated with each algorithm at each site and the combined ROC for all sites. Table III shows the ranking of each algorithm by metric at each site. Table IV shows the highest ranking algorithm by metric at each site.

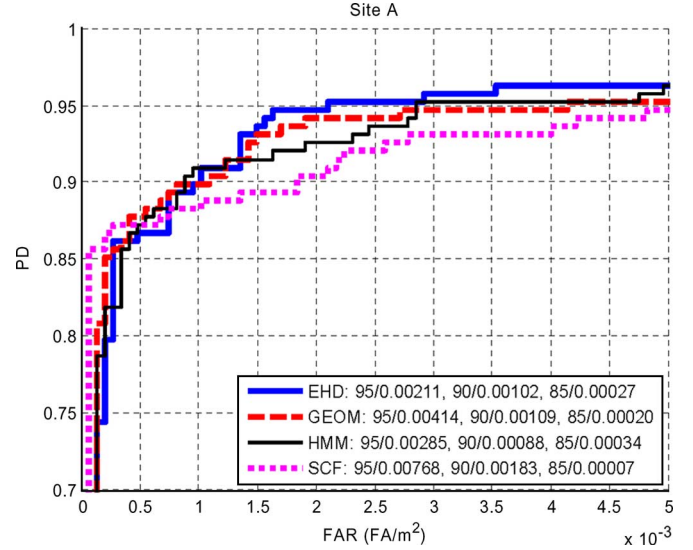


Fig. 6. Algorithm ROCs for Site A.

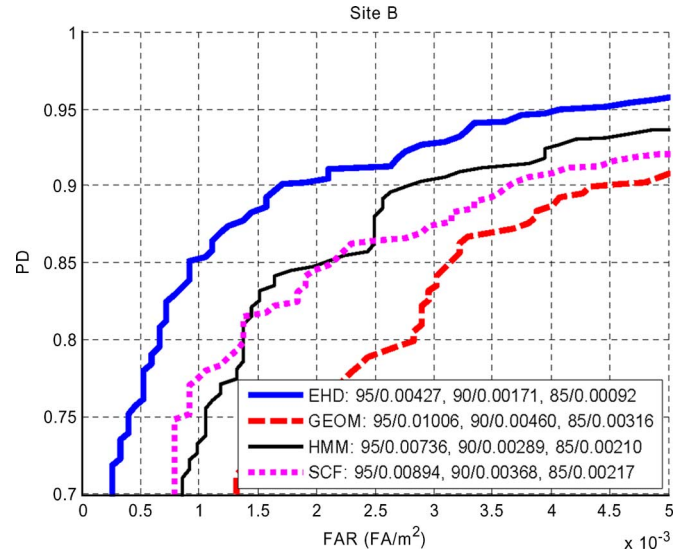


Fig. 7. Algorithm ROCs for Site B.

V. ANALYSIS AND CONCLUSION

Our goal was to evaluate a collection of landmine discrimination algorithms to determine their suitability for use in an automated detection system in a variety of different locations. We carried out an evaluation using a large set of data collected over an extended period of time in vastly different soil and weather conditions. The evaluation used a cross-validation experiment to create ROC curves and then compared a variety of properties of those ROC curves.

Our evaluation showed that the two edge-based algorithms, EHD and HMM, provided the best overall performance in the range of detection probabilities of interest on our entire multisite data collection. At a 90% probability of detection, the false alarm rate of GEOM (0.00458) is roughly double that of HMM (0.00232). The EHD algorithm was somewhat more consistent in achieving high rankings with respect to our evaluation criteria; however, the performance of the algorithms varied from site to site. In particular, the EHD algorithm

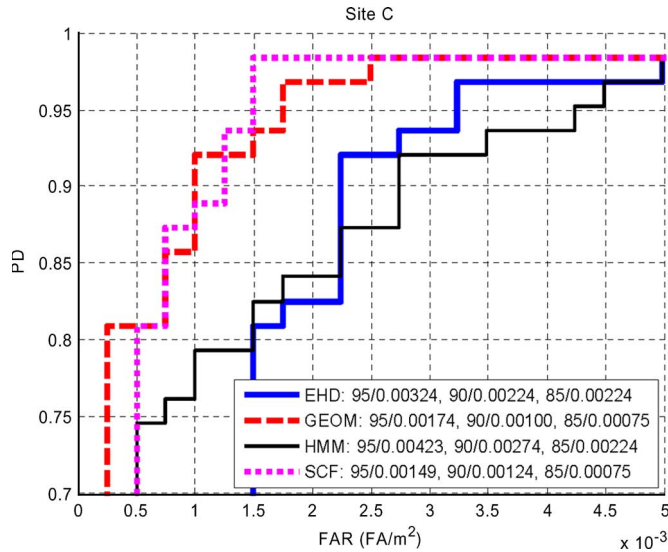


Fig. 8. Algorithm ROCs for Site C.

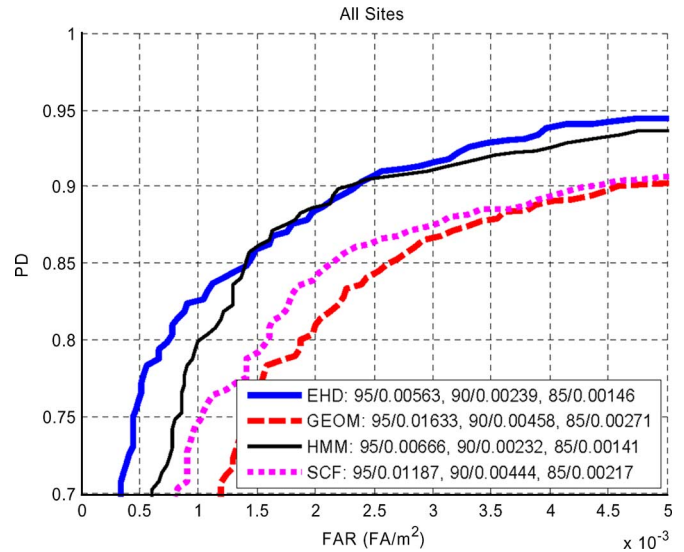


Fig. 10. Algorithm ROCs for all sites.

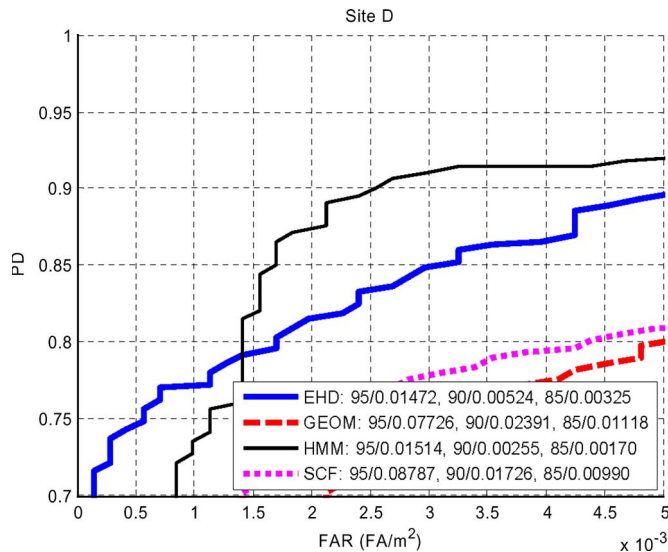


Fig. 9. Algorithm ROCs for Site D.

outperformed the HMM at PD.90 at Site B, while HMM performed better at Site D. In Fig. 7, we see that at Site B, the HMM algorithm has a larger number of false alarms from lower PDs than EHD. At this site, a single false alarm will account for $6.39e - 5$ FA/m². Thus, the difference of 0.00118 FA/m² between HMM and EHD at PD 90% is a result of about 18 false alarm occurrences. Fig. 11 shows one of the high confidence false alarms encountered by the HMM. Only a few weeks before the data shown in the figure were captured, some mines had been removed from Site B, and others had been newly laid. The soil was somewhat moist when this collection was taken. The alarm shown in Fig. 11 is reported at a location corresponding to the position of a mine that had been removed and its hole recently filled. Our conjecture is the moisture gradient between the hole and the surrounding earth accounts for this radar signature. Investigation showed that 16 of the 30 highest confidence HMM false alarms were refilled holes.

The EHD algorithm assigned these alarms much lower relative confidence than the HMM. These signatures display an edge feature sequence consistent with a buried minelike object, yet their edge histograms cluster more closely with less minelike objects.

Looking at the performance of these algorithms at Site D, we see another story. At Site D, a single false alarm corresponds to $1.368e - 4$ FA/m². Thus, the HMM algorithm has about 11 fewer false alarms than EHD at PD 85% (a difference of 0.00155 FA/m²) and 20 fewer at PD 90% (0.00269 FA/m²). In this case, it appears that the difference is a number of radar signatures in which a strong nonhyperbolic edge pattern appears, but in a sequence that is not consistent with buried minelike objects. Fig. 12 shows such an alarm in which the raw GPR signature shows a variety of edges associated with clutter, whereas the second derivative images do not show the typical hyperbolic shape we would normally associate with a buried minelike object.

Finally, the performance of both the spectral and geometric algorithms is superior at Site C than either of the HMM or EHD. In this arid soil, we see a number of mines that have extremely compact signatures displaying neither a preponderance of edges, nor clear hyperbolic features. Fig. 13 shows a typical alarm of this type. This low-metal antitank mine displays few edges in the raw GPR signal and lacks the long hyperbola tails we might normally expect to see in the second derivative images. In this case, the 0.00148-FA/m² difference between the edge algorithms and spectral/geometric algorithms at 85% PD is due to six alarms, and the 0.00124 FA/m difference is a result of five alarms at 90% PD.

These observations suggest the possibility that fusion of the algorithms results could yield a discriminator whose performance dominates all four of these algorithms.

Improvement in false alarm rates beyond the levels we have reached is difficult. Looking at the result on all sites, the best algorithm at 90% PD is the HMM. Its FAR of 0.00232 FA/m² represents 97 alarms. To achieve the goal of 0.0007 FA/m², we

TABLE III
RANKINGS OF ALGORITHM ROCs BY METRIC ON ALL SITES COLLECTION

Metric	PD95	PD90	PD85	FAR0	FAR0.0007	FAR0.00007	SEPAR
Rank1	EHD	HMM	HMM	EHD	EHD	EHD	EHD
Rank2	HMM	EHD	EHD	GEOM	GEOM	HMM	HMM
Rank3	SCF	SCF	SCF	HMM	HMM	SCF	GEOM
Rank4	GEOM	GEOM	GEOM	SCF	SCF	GEOM	SCF

TABLE IV
HIGHEST RANKING ALGORITHM BY METRIC AT EACH SITE

Metric	PD95	PD90	PD85	FAR0	FAR0.0007	FAR0.00007	SEPAR
Site A	EHD	HMM	SCF	GEOM	SCF	GEOM	GEOM
Site B	EHD	EHD	EHD	EHD	EHD	EHD	EHD
Site C	SCF	GEOM	GEOM/SCF	EHD	EHD	GEOM	SCF
Site D	EHD	HMM	HMM	EHD	EHD	EHD	HMM
All Sites	EHD	HMM	HMM	EHD	EHD	EHD	EHD

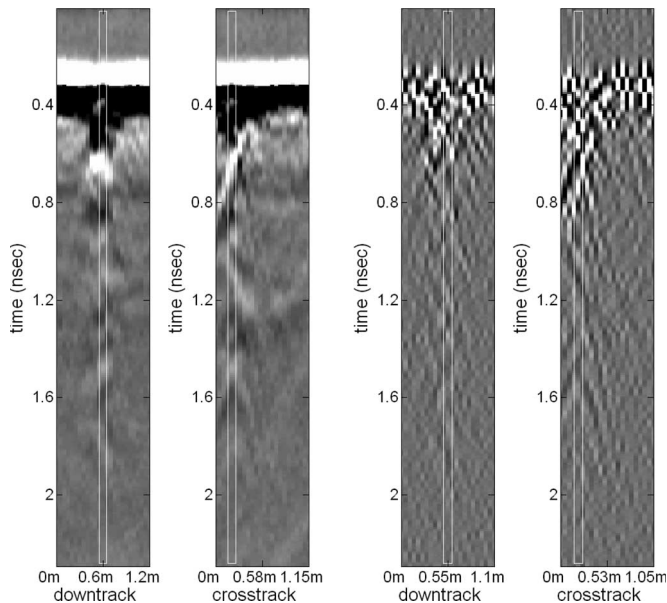


Fig. 11. B-scans of HMM high confidence false alarm at Site B. (Left) Raw GPR signature and (right) its second derivatives.

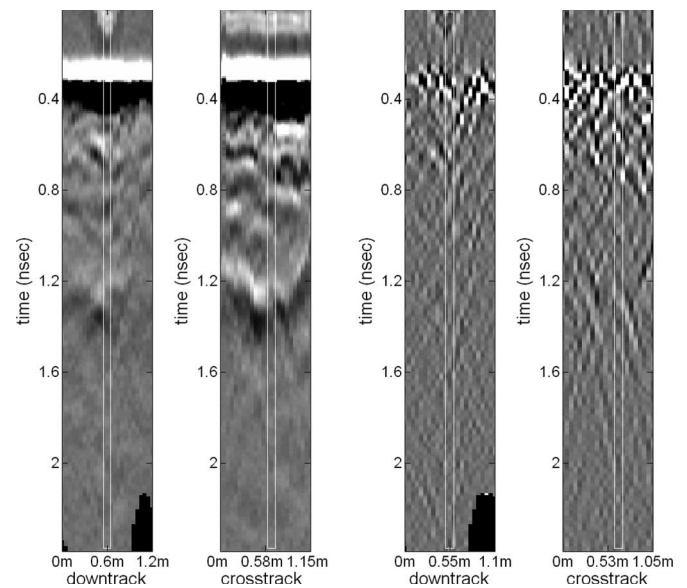


Fig. 12. B-scans of EHD high confidence false alarm at Site D. (Left) Raw GPR signature and (right) its second derivatives.

would need to reduce this to 29. There is a reason to believe, however, that such a goal may be achievable. If we present an oracle with the rank of each alarm in each algorithm in order of increasing confidence and then let it choose the highest assigned rank for each mine alarm and the lowest rank for each false alarm, then the ROC associated with this algorithm (shown in Fig. 14) has a false alarm rate of 0.00054 FA/m² at PD 90%, surpassing our goal of 0.0007. We must emphasize that such an oracle algorithm only places an upper bound on the performance of any fusion method that would use alarm ranks—because it exploits knowledge of the truth, it is not an effective algorithm. The oracle does show, however, that it is theoretically possible to fuse just the decision statistics yielded

by these algorithms, to achieve our performance goal. Our current work is oriented toward evaluating fusion algorithms for just this purpose.

ACKNOWLEDGMENT

The authors would like to thank R. Harmon, R. Weaver, P. Howard, and T. Donzelli for their support of this work, and E. Rosen and L. Ayers of IDA who provided much useful software and insight. The authors would also like to thank L. Carin, L. Collins, and P. Torriente of Duke University and F. Clodfelter and others from NIITEK, Inc., for their technical discussions, insights, cooperation, and GPR data.

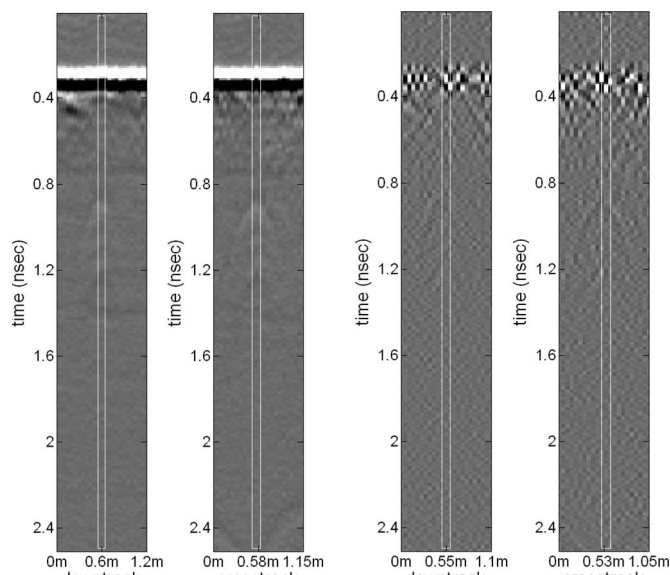


Fig. 13. B-scans of low-metal antitank mine at arid Site C. (Left) Raw GPR signature and (right) its second derivatives.

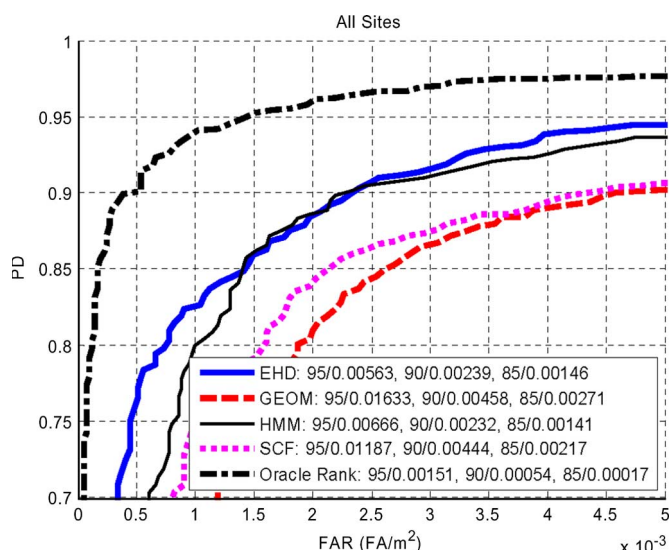


Fig. 14. Relative performance of oracle rank algorithm on all sites.

REFERENCES

- [1] C. Bruschini and B. Gros, "A survey on sensor technology for landmine detection," *J. Humanitarian Demining*, vol. 2, no. 1, Feb. 1998.
- [2] P. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1231–1244, Jun. 2001.
- [3] Y. Zhao, P. Gader, P. Chen, and Y. Zhang, "Training DHMMs of mine and clutter to minimize landmine detection errors," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pt. 1, pp. 1016–1024, May 2003.
- [4] O. Löhlein and M. Fritzsche, "Classification of GPR data for mine detection based on hidden Markov models," in *Proc. EUREL Conf. Detection Abandoned Landmines*, Edinburgh, U.K., Oct. 1998, pp. 96–100.
- [5] C. Yang, "Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 743–753, May 2005.
- [6] X. Miao, M. Azimi-Sadjadi, B. Tian, A. Dubey, and N. Witherspoon, "Detection of mines and minelike targets using principal component and neural methods," *IEEE Trans. Neural Netw.*, vol. 9, no. 3, pp. 454–463, May 1998.
- [7] B. Baertlein, "Neural-network based multi-sensor fusion for mine detection," Ballena Syst. Corp., Los Alamos, NM, Tech. Rep. N61331-93-C-0050, 1994.
- [8] S. Sheedvash and M. Azimi-Sadjadi, "Structural adaptation in neural networks with applications to land mine detection," in *Proc. IEEE Int. Conf. Neural Netw.*, 1997, pp. 1443–1447.
- [9] J. Zhang, Q. Liu, and B. Nath, "Landmine feature extraction and classification of GPR data based on SVM method," in *Proc. Int. Symp. Neural Netw.*, pt. I, Aug. 2004, pp. 636–641.
- [10] P. Torrione and L. Collins, "Application of texture feature classification methods to landmine/clutter discrimination in off-road GPR data," in *Proc. IGARSS*, 2004, pp. 1621–1624.
- [11] H. Frigui, P. Gader, and K. Satyanarayana, "Landmine detection with ground penetrating radar using fuzzy k-nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2004, vol. 3, pp. 1745–1749.
- [12] P. Gader, J. Keller, H. Frigui, H. Liu, and D. Wang, "Landmine detection using fuzzy sets with GPR images," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, May 1998, vol. 1, pp. 232–236.
- [13] H. Frigui, K. Satyanarayana, and P. Gader, "Detection of land mines using fuzzy and possibilistic membership functions," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, May 2003, vol. 2, pp. 834–839.
- [14] P. Gader, W. Lee, and A. Mendez-Vasquez, "Continuous Choquet integrals with respect to random sets with applications to landmine detection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2004, vol. 1, pp. 523–528.
- [15] P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, and A. Zare, "Multi-sensor and algorithm fusion with the Choquet integral: Applications to landmine detection," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2004, vol. 3, pp. 1605–1608.
- [16] K. J. Hintz, "SNR improvements in NIITEK ground-penetrating radar," in *Proc. SPIE, Detection and Remediation Technol. Mines and Minelike Targets IX*, Sep. 2004, vol. 5415, pp. 399–408.
- [17] Q. Zhu and L. Collins, "Application of feature extraction methods for landmine detection using the Wichmann/NIITEK ground-penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 81–85, Jan. 2005.
- [18] P. Torrione, L. Collins, F. Clodfelter, S. Frasier, and I. Starnes, "Application of the LMS algorithm to anomaly detection using the Wichmann/NIITEK ground-penetrating radar," in *Proc. SPIE, Detection and Remediation Technol. Mines and Minelike Targets VIII*, Sep. 2003, vol. 5089, pp. 1127–1136.
- [19] P. Torrione, C. S. Throckmorton, L. M. Collins, J. F. Clodfelter, S. Frasier, I. Starnes, S. S. Bishop, P. Gugino, P. Howard, R. C. Weaver, and E. Rosen, "Feature-based processing of prescreener-generated alarms for performance improvements in target identification using the NIITEK ground-penetrating radar system," in *Proc. SPIE, Detection and Remediation Technol. Mines and Minelike Targets IX*, Sep. 2004, vol. 5415, pp. 984–995.
- [20] R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue, *Survey of the State of the Art in Human Language Technology*. Washington, DC: Nat. Sci. Found., Commission Eur. Communities, Oregon Graduate Inst., 1995.
- [21] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [22] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [23] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probability functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, Feb. 1970.
- [24] L. E. Baum and T. Petrie, "Statistical inference for probability functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, no. 6, pp. 1554–1563, Dec. 1966.
- [25] H. Frigui, P. Gader, and K. C. Ho, "Real-time landmine detection with ground-penetrating radar using discriminative and adaptive hidden Markov models," *EURASIP J. Appl. Signal Process.*, vol. 12, pp. 1867–1885, 2005.
- [26] M. Mohamed and P. D. Gader, "Generalized hidden Markov models—Part II: Applications to handwritten word recognition," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 82–95, Feb. 2000.
- [27] P. D. Gader, W.-H. Lee, and J. N. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.
- [28] W.-H. Lee, P. D. Gader, and J. N. Wilson, "Optimizing the area under a receiver operating characteristic curve with application to landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 389–397, Feb. 2007.
- [29] N. Otsu, "A threshold selection method for gray-level histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, 1978.
- [30] R. Grandhi, P. D. Gader, and J. N. Wilson, "Feature analysis for the NIITEK ground-penetrating radar using order-weighted averaging

operators for landmine detection," in *Proc. SPIE, Detection and Remediation Technol. Mines and Minelike Targets IX*, Sep. 2004, vol. 5415, pp. 953–962.

- [31] R. Yager, "On order weighted averaging aggregation operators in multicriteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, Jan./Feb. 1988.
- [32] J. Fodor, J. Marichal, and M. Roubens, "Characterization of the ordered weighted averaging operators," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 2, pp. 236–240, May 1995.
- [33] R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators, Theory, and Applications*. Boston, MA: Kluwer, 1997.
- [34] H. Frigui and P. D. Gader, "Detection and discrimination of landmines in ground-penetrating radar based on edge histogram descriptors," in *Proc. SPIE, Detection and Remediation Technol. Mines and Minelike Targets X*, May 2006, vol. 6217, p. 621 733.
- [35] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG7: Multimedia Content Descriptions Language*. Hoboken, NJ: Wiley, 2002.
- [36] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Norwell, MA: Kluwer, 1981.
- [37] L. Ayers and E. Rosen, "MIDAS: Mine detection assessment and scoring user's manual V1.1," Inst. Defense Anal., Arlington, VA, Tech. Rep., 2004.
- [38] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions," in *Proc. 3rd Int. Conf. Knowl. Discovery and Data Mining*, 1997, pp. 43–48.
- [39] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [40] C. X. Ling, J. Huang, and H. Zhang, "AUC: A statistically consistent and more discriminating measure than accuracy," in *Proc. 18th IJCAI*, 2003, pp. 329–341.
- [41] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *Proc. 15th Int. Conf. Mach. Learn.*, San Francisco, CA, 1998, pp. 445–453.



Joseph N. Wilson (M'05) received the B.S. degree in applied mathematics with the emphasis on computer science from Florida State University, Tallahassee, in 1977, and the M.S. degree in applied mathematics and computer science and the Ph.D. degree in computer science from the University of Virginia, Charlottesville, in 1980 and 1985, respectively.

Since 1984, he has been a member of the faculty of the Computer and Information Science and Engineering Department, University of Florida,

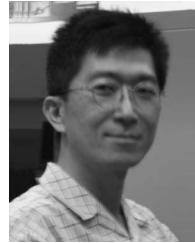
Gainesville, where he served as an Associate Chair from 1994 to 2001. His research interests include machine intelligence, image and signal processing, programming languages, and file systems.



Paul Gader (M'87–SM'99) received the Ph.D. degree in mathematics from the University of Florida, Gainesville, in 1986.

He worked as a Senior Research Scientist with Honeywell's Systems and Research Center, as a Research Engineer and Manager with the Environmental Research Institute of Michigan, and as a faculty member with the Universities of Wisconsin–Oshkosh, Missouri–Columbia, and Florida, where he is currently a Professor of computer and information science and engineering. He has led teams involved

in real-time handwritten address recognition systems for the U.S. Postal Service developing algorithms for handwritten digit recognition and segmentation, numeric field recognition, word recognition, and line segmentation and teams that devised and field-tested several real-time mine detection algorithms. He has worked on landmine detection problems involving ground-penetrating radar, electromagnetic induction, acoustic/seismic, EO/IR, and hyperspectral imaging sensors. He has over 165 technical publications in the areas of image and signal processing, applied mathematics, and pattern recognition, including over 55 refereed journal articles.



Wen-Hsiung Lee received the B.S. degree in electronics engineering from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1990, and the Master and Ph.D. degrees from the University of Florida, Gainesville, in 1994 and 2001, respectively.

From 2002 to 2006, he was a Postdoctoral Researcher with the Department of Computer and Information and Engineering, University of Florida. Since August 2006, he has been a Senior Scientist with NIITEK, Inc., Sterling, VA.



Hichem Frigui received the Ph.D. degree in computer engineering and computer science from the University of Missouri, Columbia, in 1997.

From 1998 to 2004, he was an Assistant Professor at the University of Memphis, TN. He is currently an Associate Professor and the Director of the Multimedia Research Laboratory, University of Louisville, Louisville, KY. He is currently an Associate Editor of *Fuzzy Sets and Systems*. He has published over 80 journal articles and refereed conference proceedings.

His research interests include pattern recognition, data mining, and image processing, with applications to landmine detection and content-based multimedia retrieval.

Dr. Frigui is a member of the IEEE Computer Society and ACM. He is currently an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is a recipient of the U.S. National Science Foundation Faculty Early Career Development Award for outstanding young scientists.



K. C. Ho (S'89–M'91–SM'00) was born in Hong Kong. He received the B.Sc. (with First-Class Honors) degree in electronics and the Ph.D. degree in electronic engineering in 1988 and 1991, respectively.

From 1991 to 1994, he was a Research Associate at the Royal Military College of Canada, Kingston, ON, Canada. He joined the Bell-Northern Research, Montreal, QC, Canada, in 1995 as a member of Scientific Staff. From September 1996 to August 1997, he was a member of the faculty of the Department of

Electrical Engineering, University of Saskatchewan, Saskatoon, SK, Canada. Since September 1997, he has been with the University of Missouri, Columbia, where he is currently an Associate Professor with the Electrical and Computer Engineering Department. He is also an Adjunct Associate Professor with the Royal Military College of Canada. Since 1995, he has been active in the development of the International Telecommunication Union (ITU) Standard Recommendation G.168. He is the Editor of the *ITU Standard Recommendations G.168: Digital Network Echo Cancellers* and *G.160: Voice Enhancement Device for Mobile Applications*. He is the holder or coholder of three U.S. patents, three Canadian patents, two patents in Europe, and four patents in Asia on mobile communications and signal processing. His research interests are statistical signal processing, source localization, subsurface object detection, wavelet transform, wireless communications, and the development of efficient adaptive signal processing algorithms for various applications, including landmine detection, echo cancellation, and time delay estimation.

Dr. Ho was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2006 and is currently an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He received the Junior Faculty Research Award from the College of Engineering, University of Missouri, in 2003.

Use of the Borda Count for Landmine Discriminator Fusion

J.N. Wilson, P.D. Gader

Computer and Info. Sci. and Engr. Dept., 301 CSE Bldg, Univ. of Florida, Gainesville, FL 32611

ABSTRACT

The Borda Count was proposed as a method of ranking candidates by combining the rankings assigned by multiple voters. It has been studied extensively in the context of its original use in political elections and social choice-making. It has recently seen use in machine learning and in ranking web searches, but few of its formal properties have been extensively investigated. In this paper, we describe unsupervised, and (barely) supervised learning systems that employ the Borda Count as their underlying bases. We analyze the strengths and weaknesses of the technique in the context of landmine discrimination. We discuss and evaluate methods for algorithm fusion using several weighted Borda Count approaches and show how they affect algorithm fusion performance.

Keywords: Borda Count, Landmine discrimination, fusion, unsupervised learning.

1. INTRODUCTION

This paper is concerned with combining the results of multiple algorithms for discriminating between landmines and other objects in data produced by a variety of sensors. Each discrimination algorithm is assumed to be a function of some spatially indexed data together with a value identifying a location of interest. The discriminator returns a scalar value denoting the confidence that the identified object is a landmine, greater values indicating greater confidence and lesser values indicating lesser confidence.

The current study was motivated by an observation of Michael May¹. In analyzing the landmine discrimination capabilities of 20 different algorithms operating on data from a field test of two different sensor platforms with at total of four different sensors, he calculated the sum of the ranks (by object) of the landmine confidence assigned to each object. It was noted that if one used this ranking as a discriminator, it yielded *perfect* discrimination results, that is, the landmines all had rank sums higher than the rank sums of any nonmine object. It was not immediately clear that this result could form the basis for a reasonable landmine discrimination algorithm, because it depended upon having four different sensors and the confidence results of twenty different detectors on the entirety of the data from a minefield.

Our goal in this work is to investigate the possibility of developing both supervised and semi-supervised rank-based algorithms for discriminator fusion. In this setting, we wish to use information about the ranks of various alarms in a training data set to be able to map a collection of discriminator mine confidence values into a fused mine confidence value.

2. PREVIOUS WORK

On June 16, 1770, J.C. de Borda presented a new method of election to the French Royal Academy of Sciences¹. His method involved having each voter rank all the candidates in an election. These ranks would be combined by summing, and the candidate with the best rank sum would be the winner. Soon after, the Marquis de Condorcet presented an alternate method of using pairwise comparisons to generate ranked election results. Black, Arrow, and others have analyzed the Borda, Condorcet, and other such methods for making communal ranking decisions. Each such ranking process involves a set of candidates and set of voters. The voters supply a schedule indicating their rankings (either total or pairwise) of the candidates. The following possible conditions of decision making processes based on such schedules were enumerated by Arrow:

- i. *Pairwise Comparison.* The ranking procedure makes choices between candidates in a pairwise fashion.

- ii. *Monotonicity*. If given a set of schedules yields a ranking of candidate x above y , then replacing this set of schedules with one that preserves the ordering of x and y shall yield a ranking of candidate x above y .
- iii. *Unanimity*. If in a set of schedules, x ranks above y in all schedules, x shall be ranked above y .
- iv. *Non-labeling of voters*. Interchanging the schedules of two voters shall not affect the ranking outcome.
- v. *Non-labeling of candidates*. Interchanging both the labels and ranks of two candidates on each schedule shall not affect the ranking outcome.

Arrow's impossibility theorem shows that for any procedure that satisfies all of these conditions, where there are more than 3 candidates or voters, there is at least one set of schedules that gives rise to an intransitive ranking, that is, there exist some candidates x , y , and z such that $r(x) \geq r(y)$ and $r(y) \geq r(z)$, yet $r(x) < r(z)$. The Condorcet election method, which satisfies Arrow's conditions, fails in this regard. The Borda count fails to satisfy Arrow's first (pairwise comparison) condition, but it does yield a transitive ranking result. A weighted Borda count (in which each elector's rank is multiplied by a scalar weight before summing the ranks) fails to match condition iii, however, in our setting social justice is not a concern, so this condition may be discarded. A weighting of voters is termed a *static* weighting. Likewise, condition iv can be abridged if we can determine a way in which we generate a better final confidence rank by associating different weights with different candidates. A scheme in which candidate weights may vary is termed *dynamic*.

The Borda count has been used for fusing the results of classifiers for the task of handwriting recognition.^{6,7,8} In this setting, there are C classifiers and N classes. The classes correspond to words in a lexicon. Each classifier assigns a ranking of classes (possibly partial) to each object (a handwritten word). Ho, et al., present a weighted Borda count technique for this application that uses logistic regression to identify classifier weights by comparing the ranking results of each classifier with a best ranking derived by applying several different independent classification algorithms. Gader, et al., employ a method in which the Borda weights are determined dynamically based on a match confidence between the object and a lexicon string. Van Erp and Schomaker compare the performance of the Borda count, a variant of the Borda count, in which the median rank (rather than sum or average) is used, and Nanson's election procedure (an iterative Borda scheme that deletes the candidate ranked lowest in each successive iteration).

None of these applications of the Borda count to handwriting can be applied to our fusion problem. In the handwriting case, the number of classes is large (the size of the lexicon), yet in our case the number of classes is two (mine or nonmine). Rather than generate a ranking of these two classes (in effect a decision procedure), we wish to develop a class membership confidence value for a distinguished class. Ho's use of logistic regression relies on the ability to associate a *best* ranking with a set of training instances. Although it may be possible to identify such a ranking for the class membership of handwritten words, it is not possible to identify *a-priori*, a best ranking of a set of objects all belonging to the same class (mine or nonmine). Similarly, there is no corollary to Gader's object/lexicon string match in associating a rank confidence. Van Erp and Schomaker's use of the median Borda count suggests to us the possibility of using any of a number of order-weighted averaging operators¹⁰ for static weightings but they provide no insight into how to select such a weighting scheme. The use of Nanson's procedure, however, is inappropriate for our task because, unlike normal election rankings, low confidence rankings for discrimination are no less important than high confidence rankings, thus we must not eliminate any candidates (either low or high) in performing our fusion.

3. GENERAL APPROACH

Our general approach to implementing discriminator fusion with a supervised learning system using rank weightings is to consider each discrimination algorithm to be a voter, and each alarm in the training set to be a candidate. We are given algorithms $\alpha_1, \dots, \alpha_M$ and training set sample alarm candidate objects o_1, \dots, o_N . Each algorithm maps alarms to their confidence values, elements of R . Algorithm i assigns rank $r_i(c_{ij})$ to candidate j if $c_{ij} = \alpha_i(o_j)$ has a confidence value greater than exactly $r_i(c_{ij}) - 1$ other candidate alarms. Thus, r_i is a map from the confidence values assigned by algorithm α_i into the set $\{1, \dots, N\}$. We can extend r_i to apply to a new candidate o^* with $c_i^* = \alpha_i(o^*)$ by defining

$$\hat{r}_i(c_i^*) = r_i \left(\bigvee_{c_{ij} \leq c_i^*} c_{ij} \right), \quad (3)$$

adopting the convention that the maximum of the empty set is 0. Thus, $\hat{r}_i(c_i^*)$ is the number of candidates in the training set having confidence value no greater than c_i^* .

We can now define the result of applying the (unweighted) Borda count to alarm with confidence by

$$B(o^*) = \frac{1}{MN} \sum_{i=1}^M \hat{r}_i(c_i^*). \quad (4)$$

Note that we normalize this result to yield a value in the range [0,1]. Although the algorithms may employ *a-priori* information about the training set in order to generate their confidences, the unweighted Borda fusion function B makes no use of such information. In order to determine a discriminant value, however, one must reasonably take some *a-priori* information into account. Since the confidences are generated by B in an unsupervised manner, a new alarm can be accreted to the training set by adjusting each of the rank function as follows:

$$r'_i(c_i) = \begin{cases} r_i(c_i) & \text{if } c_i < c_i^* \\ \hat{r}_i(c_i) & \text{if } c_i = c_i^* \\ r_i(c_i) + 1 & \text{if } c_i > c_i^* \end{cases} \quad (5)$$

4. SAMPLE APPLICATION

Data was collected over a grid of 220 1 meter square cells using a robot-vehicle mounted GPR sensor array and multiple passes of a single wideband metal detector. Each cell contained either a buried mine, a buried clutter object, or no buried object. Of these cells, data from 216 could be processed by all discrimination algorithms employed. Altogether, the collection contained 112 mine encounters, 64 clutter object encounters, and 40 blank cell encounters. The mines included a variety of antitank and antipersonnel mines buried at depths from 0 (flush buried) to 12.25 cm.

We processed the data with three different algorithms, one employing data from the wideband metal detector, and two employing data from the GPR sensor. The metal detector algorithm, *MD*, finds the parameters of the best fit of the data to a model proposed by Miller, et al.¹¹ and employs a two-layer, feed-forward network trained to discriminate between landmines and clutter. One of the GPR-based algorithms employs a hidden Markov model¹² (*HMM*) to discriminate between mines and clutter, and the other employs band-features of the frequency spectrum confidence feature (*SCF*) as its value.¹³

Receiver Operating Characteristic (ROC) curves were prepared as follows. The SCF algorithm is based on a simple model of mine characteristics derived from another data set and is not trained, thus, it was applied to the data associated with each object and the ROC curve was prepared in the usual way. The HMM algorithm employs models that were trained on a separate data collection from a different GPR sensor, thus it was also applied to the data associated with each object to yield a confidence value. The MD algorithm ROC, on the other hand, was prepared based on ten-way cross-validation training. The data were divided into ten test groups (each containing approximately one tenth of the objects). For each test group, the network was trained 50 times on the remaining data, and the average result of the network on the test set was used as the discriminator confidence value. The ten test groups were then cumulated together to yield the ROC curve for the entire data collection.

Figure 1 presents ROC curves yielded by application of the MD algorithm and the SCF algorithm as well as their unweighted Borda fusion. Figure 2 shows the unweighted Borda fusion of the SCF and HMM algorithms. (Note that the probability of detection axis (PD) for these graphs starts at PD 0.5.) While unweighted Borda fusion of MD and SCF yields improved performance (the fusion ROC dominates the other curves when comparing mines to blanks and dominates above PD 0.72 when applied against clutter), fusing SCF and HMM does not. On the other hand as Figure 3

shows, although unweighted Borda fusion of all three algorithms dominates the performance of the individual algorithms as a detector (against blanks), it only outperforms the MD algorithm for detection probabilities above .96 when applied to clutter.

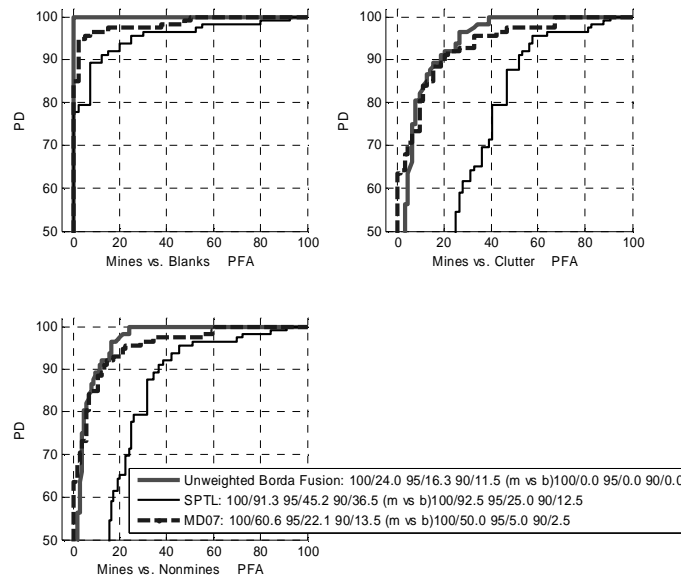


Figure 1. Unweighted Borda fusion of SCF and MD (training run of March 7). The three ROCS show performance comparing mines to blanks (upper left) mines to clutter (upper right) and mines to nonmines (lower left).

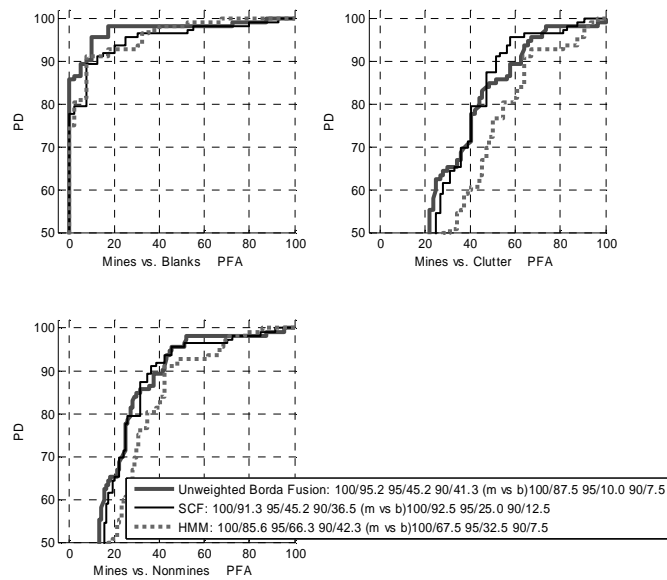


Figure 2. Unweighted Borda fusion of SCF and HMM.

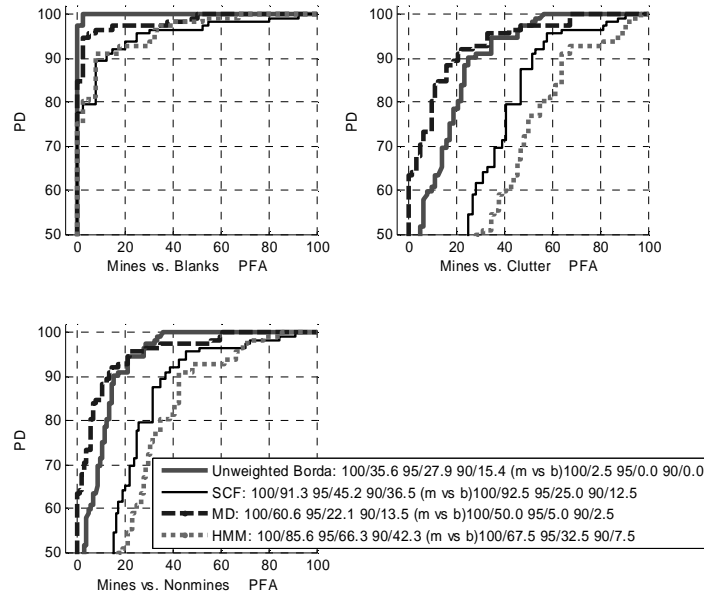


Figure 3. Unweighted Borda Fusion of SCF, MD, and HMM.

5. BORDA WEIGHTING SCHEMES

In this section, we investigate a variety of schemes to associate weights with the voters in a Borda count voting scheme.

In this setting, we define the confidence of a new alarm o^* with $c_i^* = \alpha_i(o^*)$ to be $B(o^*) = \frac{1}{MN} \sum_{i=1}^M w_i \hat{r}_i(c_i^*)$, where

w_i is the weight assigned to algorithm i . We begin this investigation by looking at ways to compare the similarity of rankings.

Kendall defined the *rank correlation coefficient*¹⁵, τ , which is a measure of the similarity of two rankings. This coefficient can be defined on two rankings r and s of objects as follows:

$$\tau(r, s) = \frac{2}{n(n-1)} \sum_{i=1}^N \sum_{j \neq i} \text{sgn}(r(i) - r(j)) \text{sgn}(s(i) - s(j)) \quad (6)$$

Thus τ is the normalized sum of the number of agreements in ordering of pairs of items minus the number of disagreements in pairs of orderings. The value of τ varies between -1 (for exactly opposite rankings) and 1 (for identical rankings).

Kendall also defined the *coefficient of concordance* of $m > 1$ rankings, W , defined for m rankings $R = \{r_1, \dots, r_m\}$ of n objects as follows:

$$W(R) = \frac{12}{m^2(n^3 - n)} \sum_{j=1}^n \left(\left(\sum_{i=1}^m r_i(j) \right) - \frac{1}{2} m(n+1) \right)^2. \quad (7)$$

This sums the normalized deviation of the sum of the ranks of an object from the mean sum of ranks over all objects. These coefficients have been employed by several authors in connection with Borda rank fusion methods. Erp and Schomaker¹¹ employ W to compare rankings between Borda's algorithm, a median-weighted Borda count, and Nanson's algorithm. Sculley uses τ to evaluate the performance of several ranking algorithms including a weighted Borda count, by comparing the ranking yielded by the algorithm to a *best* ranking. As noted above, no definitive best ranking of discrimination confidences can be identified. However, we can define a discrimination confidence ranking to be *accurate* if a higher confidence rank is associated with each mine object than is associated with any nonmine object.

One may well ask whether these coefficients themselves might be gainfully employed to identify Borda weightings. We conjecture that for a given ranking of mine and nonmine objects r , that the accurate ranking most highly correlated with r , namely, r' , is that which preserves the orderings of the mines and nonmines, but ranks all mines higher than nonmines. It is reasonable to say that algorithm i is a better discriminator than algorithm j if $\tau(r_i, r'_i) > \tau(r_j, r'_j)$. Indeed, for the data referred to earlier, we find that these correlation coefficients are the following:

$$\begin{aligned}\tau(\text{MD}, \text{MD}') &= .952, \\ \tau(\text{SCF}, \text{SCF}') &= .829, \\ \tau(\text{HMM}, \text{HMM}') &= .792,\end{aligned}$$

which is not surprising because the ROC for MD dominates the ROC for SCF, and the SCF ROC dominates the HMM ROC until the probability of detection approaches 1. One might be tempted to use these correlation values as weights, however, application to our sample as shown in Figure 4 demonstrates that this may not yield a dominating ROC.

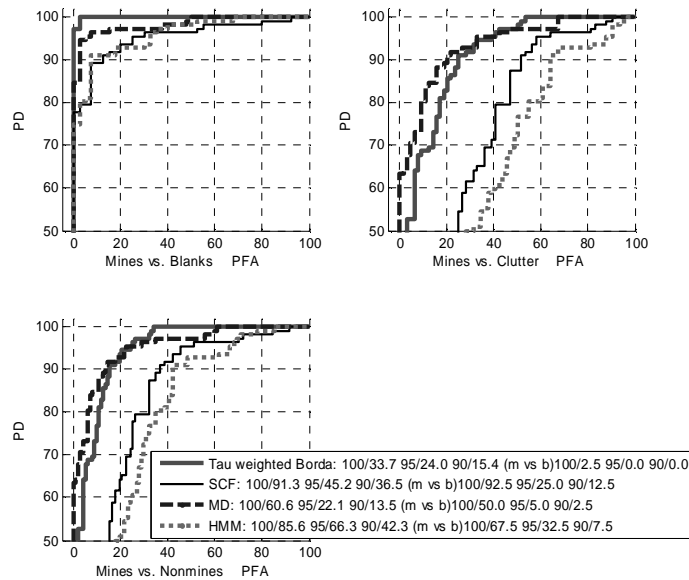


Figure 4. Tau-weighted Borda fusion.

One might argue that the weights associated with the different algorithms are too low, because the range of values of τ is from -1 to 1. On the other hand, one must consider several practical issues. There is no possibility of this process yielding a τ value of -1 because the order of mine (resp. nonmine) value pairs is not changed by the process. Furthermore, the worst out of order case would be a ranking in which all nonmines are ranked with confidences below all mines. In such a case, one has a perfectly accurate discriminator with confidence weighting the mines lower than nonmines rather than higher. In fact the worst case ROC for a detector is the chance diagonal, which corresponds to an interleaving of mines and nonmines in the ranking. Experimental evidence indicates that $\tau(r, r') \rightarrow 0.5$ as the number of objects $n \rightarrow \infty$ if r has a chance diagonal ROC. Thus, we might consider using $\tau(r, r') - 0.5$ as a better algorithm

weight estimator. Figure 5 shows that this approach, while better than the previous τ weighting, still does not yield a dominating ROC.

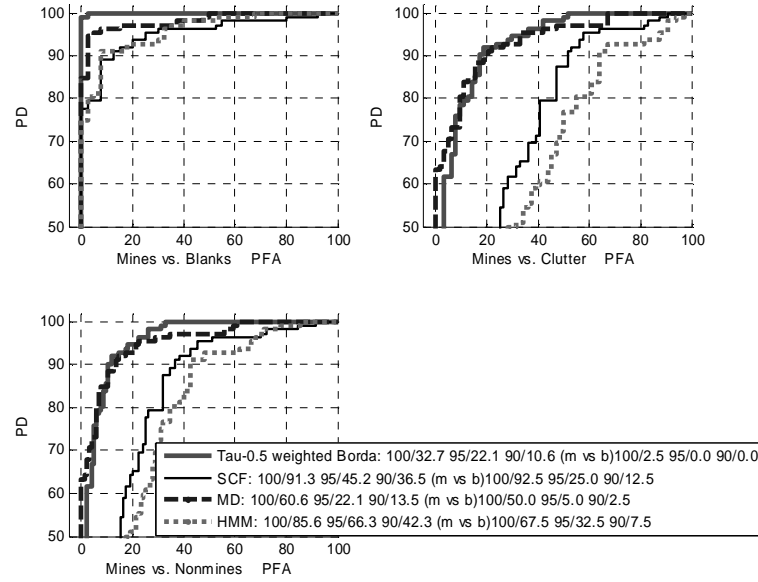


Figure 5. Borda weighted using $\tau=0.5$.

One can apply the theory of gambling to the problem of assigning weights to different discriminators. A discussion of the relationship between information theory and gambling theory is contained in Cover and Thomas¹¹. Consider that each discriminator $i \in \{1, \dots, M\}$ represents a participant in a number of contests corresponding to each alarm object $n \in \{1, \dots, N\}$. The payoff for a win by contestant i , that is, the number of dollars returned for a one dollar bet if contestant i wins, is \hat{o}_i . Let b_i represent the bet (fraction of wealth) wagered on contestant i satisfying $b_i \geq 0$ for all i and $1 = \sum_{i=1}^M b_i$. Let p_i represent the probability that contestant i will win the contest. The *doubling rate* (the fraction by which each contest will yield a doubling of wealth) is given by

$$W(b, p) = \sum_{i=1}^M p_i \log b_i \hat{o}_i. \quad (1)$$

It has been shown that under these conditions, the optimal (largest) doubling rate is given by

$$W^*(p) = \sum_{i=1}^M p_i \log \hat{o}_i - H(p) \quad (2)$$

where $H(p)$ represents the entropy of p , and is given by $b^* = p$, that is, each bet proportion b_i should match the win proportion p_i of each contestant.

To apply this theory appropriately, we must be able to determine the distribution p , which gives the probability that a discriminator will win a ranking contest with other discriminators, and this is no easy matter. One might consider looking at the number of objects for which each algorithm's rank is the best (i.e., highest rank value for a mine and lowest for a nonmine) however, application of this concept in our sample data yields ROCs shown in Figure 6, and once

again, a dominating ROC is not found. It might be possible to better define what it means for one algorithm to win a ranking contest for a given alarm, however, it might be difficult to do this without appealing to the notion of a best rank for each item, which notion we have already rejected as described above.

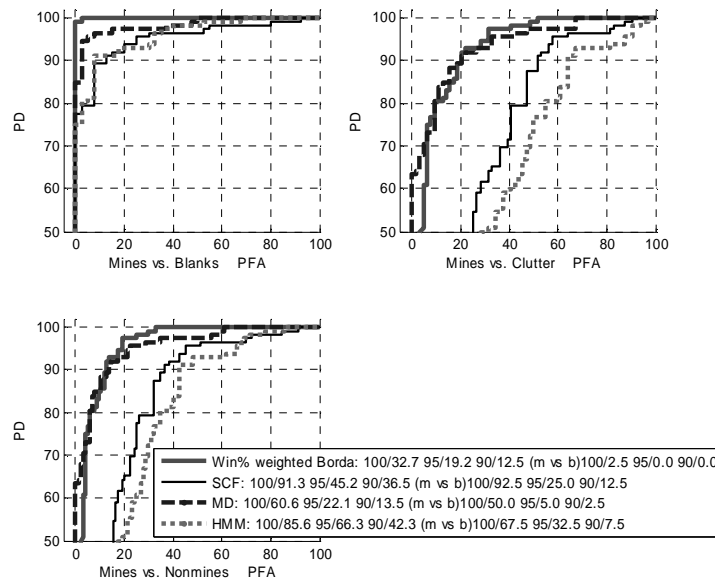


Figure 6. Borda weighted by probability of best alarm rank.

Our final approach to this problem is to address our actual evaluation criterion, namely the ROC curve itself. Recent work¹¹ has shown that the area under the ROC curve (AUC) is an unbiased estimator of discrimination accuracy and ROC curve area optimization is not a new concept. We can apply AUC optimization to the problem of generating Borda weights by exhaustively searching the weight space to find those that yield the largest AUC. We applied this approach to our sample data set. Figure 7 shows the ROC area as a function of the weights of SCF and HMM for one crossvalidation fold.

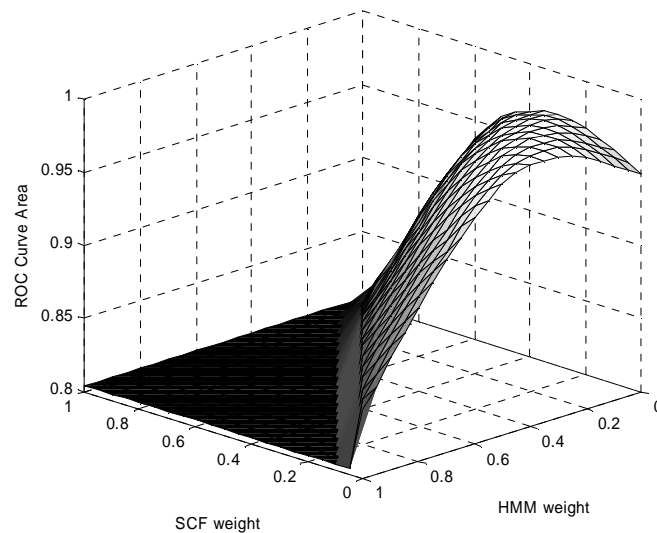


Figure 7. ROC Area as a function of SCF and HMM weight.

Although only one crossvalidation fold is shown, each of the folds yields a similar curve with maximal value in a roughly linear region in which the sum of the SCF and HMM weights are around .3 to .35. In the cross-validation run shown, the weights of the HMM ranged from 0 to .15 and the SCF weights from .2 to .3. Employing the weights identified by maximal AUC search yields the ROC curves of Figure 8. The AUC-weighted Borda dominates the other ROCs above a detection probability of about .65, and yields the lowest false alarm probability weights of any of the methods used above.

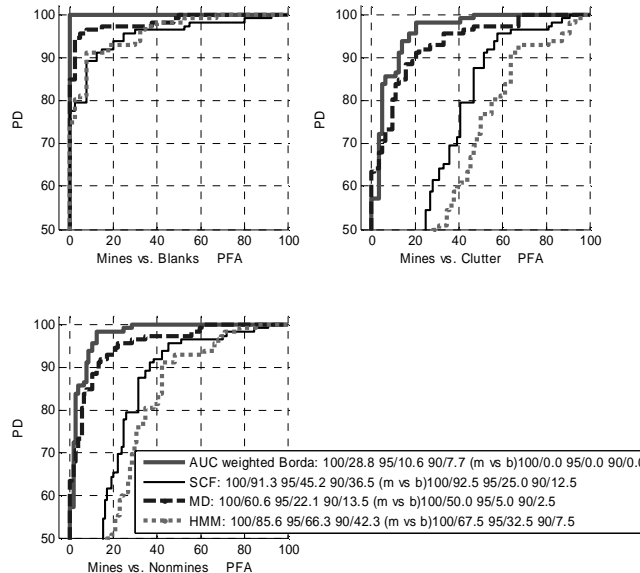


Figure 8. Maximum AUC weighted Borda fusion.

The above techniques have used static algorithm weightings to do their job, that is, each algorithm is assigned a single weight for all objects to be ranked. One can employ dynamic ranks that depend on some function of the object ranks or some other independent information derived from the object data. Figure 9 shows the results that would be achieved by a cross-validation oracle, one that could somehow determine for each algorithm object j , the algorithm i that yields the best cross-validation set rank for that object (the highest rank for a mine object and the lowest rank for a nonmine), and assigns weight 1 to algorithm i for that alarm and weight 0 to all other algorithms. Thus, the ROCs shown are still derived using cross-validation, but require an oracle to identify which algorithm works best for that alarm. The oracle's performance is stellar, yielding a practically perfect ROC curve. On the other hand, using dynamic weighting could potentially produce very bad results. Figure 10 shows the results of consulting an oracle who lies about the disposition of each alarm, setting the weights to assign the worst possible algorithm's rank. This algorithm's poor performance indicates we should use caution when considering the use of dynamic weights.

6. CONCLUSION

We have investigated the application of the Borda count to the fusion of discriminator confidence values. We showed the properties of the Borda count that make it more suitable to this task than either Condorcet's or Nance's voting procedures. We briefly reviews some of the applications of the Borda count to the problem of handwriting recognition and identified those properties of the problem domain that differ from the problem of discriminator fusion. We addressed several different methods of assigning weights to discriminators based on rank correlation and gambling

theory and found them to yield minor improvements in discrimination capability as shown by ROC curves. We employed exhaustive search over the space of discriminator weightings using area under the ROC curve as our optimization criterion, and achieved improved ROC curves. Finally, we should how one can bound the results of dynamically weighted Borda fusion using oracles on the training set, and observed that though dynamic weighting strategies may have the potential to provide much better performance, they may be subject to dramatic failure as well.

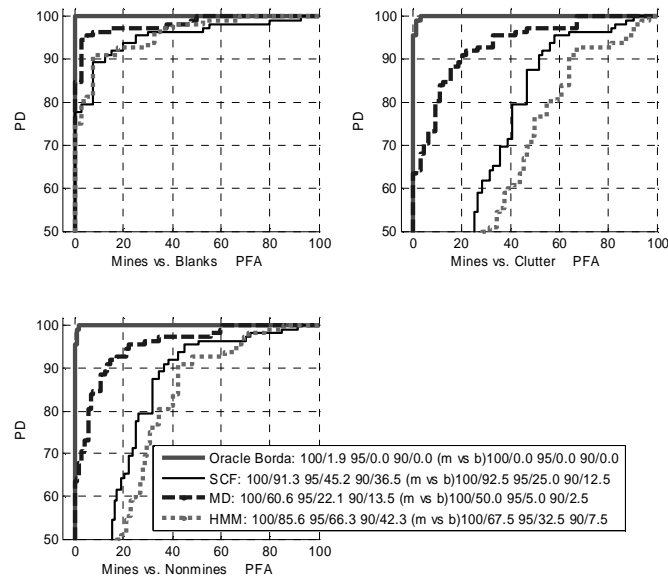


Figure 9. Oracle Borda ranking.

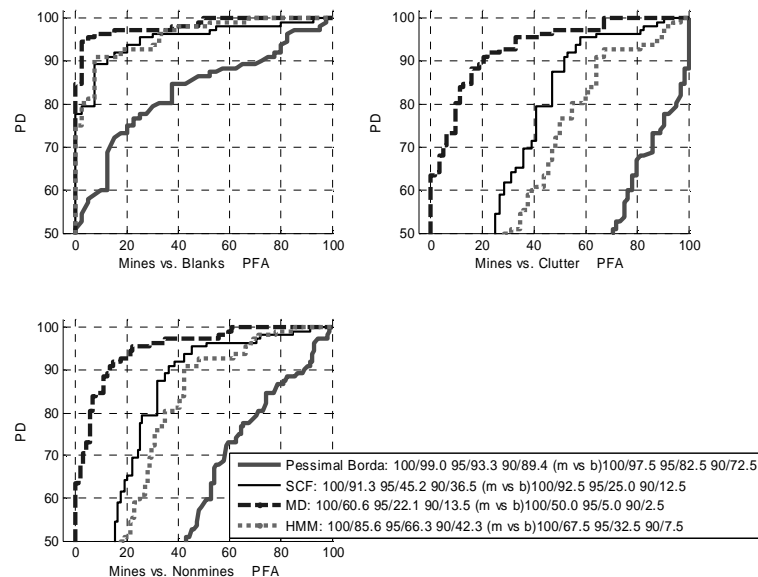


Figure 10. Pessimistic oracle Borda ranking.

7. ACKNOWLEDGEMENTS

This research was partially supported by Army Research Office grant #W911NF-05-1-0067. We thank Russell Harmon, Richard Weaver, Eloisa Lara, Mark Locke, Frank Navish, and Brian DeGrano for their support of this work. Michael May of IDA provided the initial motivation for this work. We also thank Fred Clodfelter and others from NIITEK, Inc. as well as Waymond Scott from GTRI for their technical discussions, insights, cooperation, and data.

REFERENCES

1. May, M., *personal communication*.
2. de Borda, J.C., "Mémoire sur les élections au scrutin," *Histoire de l'Académie Royale des Sciences*, Paris, 1781.
3. Condorcet, M.J.A.N. de Caritat, Marquis de, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris, Imprimerie Royale, 1785.
4. Black, D., *The Theory of Committees and Elections*, 2 Ed, Kluwer Academic Publishers, Boston, 1998.
5. Arrow, K.J., *Social Choice and Individual Values*, 2 Ed, Yale University Press, New Haven, CT, 1963.
6. Ho, T.K., Hull, J.J., and Srihari, S.N., "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1), Jan. 1994, pp 66-75.
7. Gader, P.D., Whalen, Mohamed, M.A., Keller, J.M., "Fusion of handwritten word classifiers," *Pattern Recognition Letters*, 17(6), Jun. 1996, pp 577-584.
8. van Erp, K. and Schomaker, L., "Variants of the Borda count method for combining ranked classifier hypotheses," *Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition*, Sep. 11-13, 2000, pp. 443-452.
9. McLean, I., "E.J. Nanson, social choice, and electoral reform," *Australian Journal of Political Science*, 31(3), Nov. 1996, pp 369-385.
10. Yager, R.R., "On order weighted averaging aggregation operators in multicriteria decision making," *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1), Jan. 1988, pp 183-190.
11. Miller, J.T., Thomas, H.B., Soukup, J., "Simple phenomenological models for wideband frequency-domain electromagnetic induction," *IEEE Transactions on Geoscience and Remote Sensing*, 39(6), June 2001, pp1294-1298.
12. Gader, P.D., Mystkowski, M., and Zhao, Y., "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Transactions on Geoscience and Remote Sensing*, 39(6), Jun. 2001, pp 1231-1244.
13. Ho, K.C., Gader, P.D. and Wilson, J.N., "Improving landmine detection using frequency domain features from ground penetrating radar," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Sep. 2004, 3, pp 1617-1620.
14. Peterson, W.W., Birdsall, T.G. and Fox, W.C., "The theory of signal detectability," *IEEE Transactions on Information Theory*, 4(4), Sep. 1954, pp 171-212.
15. Kendall, M.G. "A new measure of rank correlation," *Biometrika*, 30(1/2), Jun. 1938, pp 81-93.
16. Kendall, M.G. and Babington Smith, B., "The problem of m rankings," *Annals of Mathematical Statistics*, 10(3), Sep. 1939, pp. 275-287.
17. Sculley, D., *Rank Aggregation for Similar Items*, Tufts University, Technical Report, TR-2007-1.
18. Cover, T. and Thomas, J., *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
19. Ling, C.X., Huang, J. and Zhang, H., "AUC: a statistically consistent and more discriminating measure than accuracy," *International Joint Conference on Artificial Intelligence*, 18, 2003, pp 519-516.
20. Cortes, C. and Mohri, M., "AUC optimization vs. error rate minimization," *Advances in Neural Information Processing Systems*, 2004.